Ecological Modelling 388 (2018) 136–144

Contents lists available at ScienceDirect

Ecological Modelling

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

Modelling the potential distribution of an invasive mosquito species: comparative evaluation of four machine learning methods and their combinations

Linus Früh[a,d,*], Helge Kampen[b], Antje Kerkow[a,c], Günter A. Schaub[d], Doreen Walther[a], Ralf Wieland[a]

[a] *Leibniz Centre for Agricultural Landscape Research, Eberswalder Str. 84, 15374 Müncheberg, Germany*
[b] *Friedrich-Loeffler-Institut, Federal Research Institute for Animal Health, Südufer 10, 17493 Greifswald, Insel Riems, Germany*
[c] *Freie Universität Berlin, Königin-Luise-Str. 1-3, 14195 Berlin, Germany*
[d] *Ruhr-Universität Bochum, Universitätsstraße 150, 44801 Bochum, Germany*

A R T I C L E I N F O

Keywords:
Decision tree
Logistic regression
Random forest
Support vector machine
Hasse diagram technique
Aedes japonicus japonicus

A B S T R A C T

We tested four machine learning methods for their performance in the classification of mosquito species occurrence related to weather variables: support vector machine, random forest, logistic regression and decision tree. The objective was to find a method which showed the most accurate model for the prediction of the potential geographical distribution of *Aedes japonicus japonicus*, an invasive mosquito species in Germany.

The evaluation of the model trainings was conducted using derivations of a confusion matrix. Furthermore, we introduced two quality indices, 'selectivity' and 'exactness', for the evaluation of the spatial simulation, visualised through the Hasse diagram technique.

From the evaluation results we can conclude that a specific combination of two to three models performs better in predicting the potential distribution of the mosquito species than a single model or the random combination of models.

1. Introduction

Over the last two decades, the use of machine learning methods has become standard for the classification of species distribution. A lot of algorithms are available for modelling species distribution, and more models are continuously being developed (Guisan and Zimmermann, 2000; Fernández-Delgado et al., 2014). Against this background, the choice of the most accurate method can be challenging.

Modelling the distribution of an invasive species is a sensitive issue, especially when it is a potential vector of human pathogens (Fischer et al., 2014). Therefore, it is important to evaluate and compare the output of various modelling methods to learn which one produces the best results (Hastie et al., 2009). From a biological perspective, it is also necessary to check how different models predict the distribution of populations on a regional scale, i.e. if they are able to show regions through a spatial simulation that are more or less suitable for colonisation than others.

By now, a lot of different machine learning methods are applicable to biological questions (e.g. Tarca et al., 2007; Olden et al., 2008; Wieland and Mirschel, 2008; Kampichler et al., 2010; Fernández-Delgado et al., 2014). Recently, the combination of several modelling methods has become popular when classifying species distribution (e.g. Thomson et al., 2006; Grenouillet et al., 2011; Solazzo and Galmarini, 2014; Martre et al., 2015). Modelling climatically dependent potential distribution is lately done for various invasive mosquito species, e.g. *Aedes albopictus* and *Ae. japonicus japonicus*, showing potentially suitable habitats (Fischer et al., 2014; Kraemer et al., 2015; Melaun et al., 2015; Cunze et al., 2016). *Aedes j. japonicus* is an invasive mosquito species which has entered the USA and Europe through globalised trade from eastern Asia (Kampen and Werner, 2014). Monitoring in Germany showed occurrences in Baden-Wuerttemberg (Huber et al., 2012), North Rhine-Westphalia/Rhineland-Palatinate, Lower Saxony/North Rhine-Westphalia (Kampen et al., 2012; Werner and Kampen, 2013) and Bavaria (Zielke et al., 2016). The potential of the species as a vector of viruses (Takashima and Rosen, 1989; Sardelis et al., 2002a, 2002b, 2003; Schaffner et al., 2011; Huber et al., 2014) requires thorough surveillance.

Evaluating the training results of machine learning methods is

* Corresponding author at: Leibniz Centre for Agricultural Landscape Research, Eberswalder Str. 84, 15374 Müncheberg, Germany.
E-mail address: linus.frueh@zalf.de (L. Früh).

https://doi.org/10.1016/j.ecolmodel.2018.08.011
Received 8 November 2017; Received in revised form 5 June 2018; Accepted 18 August 2018
Available online 11 September 2018

usually done through performance indices, e.g. 'precision' and 'recall', which are derived from a confusion matrix (Buckland and Elston, 1993; Lobo et al., 2008; Bennett et al., 2013). Another approach to test the predictive power of a model can be done by analysing the spatial simulation of the occurrence of the species. The resulting performance indices consist in a partially ordered set (Halfon, 1985) which can be visualised using the Hasse diagram technique (HDT) (Hasse, 1952; Brüggemann et al., 1994). The HDT has been proven useful in many environmental studies, such as ecotoxicological assessments of pollutants and comparisons of land use strategies (Brüggemann et al., 1995; Voigt et al., 2006; Tsonkova et al., 2015).

Some models only work well with a certain dataset while others produce overly complex, overfitted models with default settings (Rodda et al., 2011). For the prediction of the potential distribution of *Ae. j. japonicus* it is unknown which model or combination of models is the most suitable. The previous ensemble forecasting method for the prediction of the potential distribution of *Ae. j. japonicus* in Europe used a poor dataset and without differentiating between the step-by-step combination of models (Cunze et al., 2016). Instead of this combination of all used models, a combination of selected methods should be tested (Solazzo and Galmarini, 2014; Martre et al., 2015). This study is part of a joint research project and aims to test the suitability of different models for a specific dataset of mosquito occurrence points. One of the goals of the joint research project is to provide maps of occurrence and dispersal of *Ae. j. japonicus* in Germany, based on ecological and geological factors. The primary objective of the present study is to find the model with the most accurate prediction of the potential distribution of *Ae. j. japonicus* in Germany related to weather variables, as these seem to be most important in modelling the occurrence of an invasive mosquito species (Fischer et al., 2011, 2014, Cunze et al., 2016).

For ranking the performance, the indices 'recall', 'precision', 'selectivity', 'exactness', time for training, adjustability and comprehensibility were compared. Through specific combinations of models, an improvement of the performance is expected to be achieved (Martre et al., 2015). The application of the HDT should clarify the order of the evaluation output of the single models and their combinations, respectively; i.e. the HDT should show if a single model or a combination of certain models yields the best results for predicting the potential distribution of *Ae. j. japonicus.*

2. Methods

2.1. Selection of predictors

We accessed mosquito collection data via the German national mosquito database 'Culbase', which included data from the citizen science project "Mueckenatlas" (Werner et al., 2014; Walther and Kampen, 2017) and from active monitoring activities (Kampen et al., 2016). Collection data of *Ae. j. japonicus* and three native mosquito species were extracted from the 'Culbase' database.

We followed the approach of Kerkow et al. (unpublished), who assumed that environmental variables favour one species more than others. Thus, we were able to perform a classification of mosquito species presence, including weather variables as predictors, with the following assumption, which was based on the probability of the occurrence of a mosquito species (p) as a function (f) of weather (w) (Eq. (1)):

$$p = f(w) \tag{1}$$

An analysis of publicly available weather variables in adequate resolution (1 km^2; Deutscher Wetterdienst, https://cdc.dwd.de/) according to the method by Wieland et al. (2017) to correlate with the potential occurrence of *Ae. j. japonicus* in Germany in 2014, produced the following variables to be most predictive: mean precipitation in February, April and June; mean temperature in September, October and December; mean temperature in March, April and May, and drought index in September, October and November.

Handling the dataset was done by the Python modules 'pandas' and 'numpy'.

2.2. Model selection

From the many models available we focused on those which have been evaluated with the highest ranks by Fernández-Delgado et al. (2014). Some classifier families have been excluded from the preselection because they produce better results in a different application area, e.g. neural networks are more suitable for deep learning (Kampichler et al., 2010). Other classifiers had to be rejected due to poor results in pre-tests (e.g. AdaBoost). The following models were selected: decision tree (DT), which is easy to comprehend and to interpret (Breiman et al., 1984), logistic regression (LR), a very popular and fast classifier (Cox, 1958), random forest (RF), the most powerful classifier at the moment (Breiman, 2001; Fernández-Delgado et al., 2014), and support vector machine (SVM), which works well even with small datasets (Cortes and Vapnik, 1995).

2.3. Modelling

The dataset of mosquito collection points was split into training data (years 2011–2014) and test data (year 2015). We used data of the years 2011–2014 (n = 2988), with a random selection of max. 1000 sampling points for each training step and weather data from 2014.

The Python module 'scikit-learn' integrated a bunch of modern machine learning algorithms that could be used for fast and efficient supervised classification (Pedregosa et al., 2011). It was accompanied by a precise and comprehensive documentation (Garreta and Moncecchi, 2013). We chose the four supervised algorithms from the models available in the 'scikit-learn' module (see Table 1).

For each model, we changed the default settings (e.g. the depth of the tree for DT; see Table 1) to fit the model to our data until the score of the confusion matrix could not be improved anymore. Each model passed ten training steps to decrease the risk of choosing an unrepresentative selection of the dataset. Visualisation of the results with boxplots was implemented via the Python module 'matplotlib'. To avoid

Table 1
Name, settings, originators and application examples of the selected classifiers. The specified settings improved the evaluation results, compared to results gained through default settings.

Model	Classifier	Settings	Originator	Application examples
DT	DecisionTreeClassifier	criterion = 'entropy', max_depth = 5	Breiman et al. (1984)	Podgorelec et al. (2002), Kampichler et al. (2010)
LR	LogisticRegression	C = 1.0, penalty = 'l2', tol = 0.01	Cox (1958)	Hosmer and Lemeshow (2000), Fischer et al. (2011)
RF	RandomForestClassifier	min_samples_split = 10, random_state = 0	Breiman (2001)	Pino-Mejías et al. (2010), Vezza et al. (2015)
SVM	svm.SVC	gamma = 0.00008, tol = 1e-10, probability = True	Cortes and Vapnik (1995)	Kampichler et al. (2010), Pino-Mejías et al. (2010)

retraining the models in case of future applications, we used the Python built-in persistence module ‘pickle’.

An averaging multi-model approach was used by combining the four models. The binomial coefficient could calculate the number of possible combinations (Eq. (2)):

$$\sum_{k=1}^{n}\binom{n}{k} = \sum_{k=1}^{n}\left(\frac{n!}{k!\bullet(n-k)!}\right) = 2^{n}-1 \tag{2}$$

By subtracting the four single models, we got 11 combinations.

For spatial simulation, the potential distribution of *Ae. j. japonicus* was compiled in an ASCII grid file and visualised in maps through ‘Spatial Analysis and Modeling Tool Version 2’ (SAMT2, Wieland et al., 2006, 2015) and QGIS 2.16.3. Ensemble models were produced by adding the output of specific single models and dividing the result by the number of the models in the ensemble through SAMT2.

2.4. Evaluation

The evaluation of the training was done with sampling data from 2015.

2.4.1. Training

For statistical evaluation of the single models, we took the confusion matrix (Buckland and Elston, 1993) from the training and used the Python ‘metrics’ module to show a report with the performance indices ‘precision’ and ‘recall’. We considered collection points of *Ae. j. japonicus* as a positive condition and handled the exclusive occurrence of other species as a negative condition for *Ae. j. japonicus*.

2.4.2. Spatial simulation

Due to the lack of absence data we decided to develop an own evaluation measure to quantify the prediction performance of the spatial simulation. We assumed that the lowest 10% of the predicted values can be subtracted from the prediction map, since they must be expected to represent climatically unsuitable locations, resulting in instable populations (if they provided habitable areas at all). The model should be able to predict regions with low probabilities of potential distribution. Predicting large areas with high occurrence probabilities resulting from false positives makes a model vague and unreliable. For this reason we introduced the index ‘selectivity’ (S) as a measure to explain the robustness of the model against false predictions of occurrence. It was calculated by using the collection data from 2015, taking into account the predicted probability of each point. The lowest 10% of these points were subtracted from the area of Germany ($A0$), giving a smaller area where the predicted probability (p) was greater than 10% of the lowest values of all points: $Area(p > 10\%)$. Afterwards, the index was calculated by the following equation (Eq. (3)):

$$S = \frac{(A0 - Area(p > 10\%))}{A0} \tag{3}$$

The smaller the $Area(p > 10\%)$, i.e. the higher the index S, the more selective is the model, which means that the model is able to predict accurate areas of occurrence of *Ae. j. japonicus* and to minimise false predictions of occurrence of the species.

The index E shows the power of the models to predict the occurrence of *Ae. j. japonicus* with a certain probability, i.e. the ‘exactness’ (Eq. (4)). It was calculated by the $Mean\ (p|m)$, which is the arithmetic mean of the predicted probability (p) at the collection sites of mosquitoes in 2015 (m), with exclusive consideration of collection points where the predicted probability values (p) were greater than 10%.

$$E = Mean\ (p|m)\ \wedge p > 10\% \tag{4}$$

The higher the mean of the matched collection point values of 2015 is, the more useful the model demonstrates to be for the prediction of *Ae. j. japonicus*. Therefore, the best model evaluated through the index E calculates the highest values of probability of occurrence for the collection point values of 2015 (i.e. the test data).

Both indices, S and E, are equally important, making a direct comparison inadequate. They represent a partially ordered set (Reggiani and Marchetti, 1975; Halfon, 1985; Brüggemann and Steinberg, 1998) which can be graphically visualised by the HDT (Hasse, 1952; Halfon, 1985; Brüggemann et al., 1994). The HDT is available via the PyHasse software package (https://pyhasse.org/).

In this study, the HDT was used to compare the indices (S_1, E_1) of one model (M_1) with the indices (S_2, E_2) of another model (M_2). M_1 has higher values than M_2, if S_1 is bigger than S_2 or E_1 is bigger than E_2 (Eq. (5)). M_1 and M_2 are not comparable, when S_1 is bigger than S_2 but E_1 is lower than E_2, or vice versa (Eq. (6)).

$$M_1(S_1\,E_1) > M_2(S_2\,E_2)\ \Leftrightarrow\ S_1 > S_2\ \wedge\ E_1 > E_2 \tag{5}$$

$$M_1(S_1\,E_1) \neq M_2(S_2\,E_2)\ \Leftrightarrow\ S_1 > S_2\ \wedge\ E_1 < E_2\ \vee\ \ S_1 < S_2\ \wedge\ E_1 > E_2 \tag{6}$$

2.4.3. Secondary performance indices

The time for training can vary highly between the algorithms, especially depending on the size of the dataset and the number of estimators and input variables. We measured the time needed via the Python module ‘timeit’.

If the model has a lot of adjustable options, a good performance even for a special dataset is evident. On the other side, it can be very time-consuming to find the best configuration when there are too many options. Therefore, we changed the settings of the models according to the documentation (http://scikit-learn.org, Pedregosa et al., 2011; Garreta and Moncecchi, 2013) until the results of the training could not be improved anymore.

A good comprehensibility of the model enables to retrace the decision path of the estimators. This allows to evaluate the importance of the used input variables (Pedregosa et al., 2011). The decision if the model was either a black box or a white box method was made by examining their functionality (Cox, 1958; Breiman et al., 1984; Cortes and Vapnik, 1995; Breiman, 2001).

3. Results

3.1. Performance indices

The values of the performance indices derive from 10 steps of training and testing the single models. The model with the highest values of ‘precision’, ranging from 0.8 to 0.85 and, thus, showing only small variance, is SVM. The median of RF reaches the second highest values of 0.74, but the range of the testing results of the model has a rather high variance from 0.64 to 0.81. LR has a median precision of 0.72, fluctuating from 0.67 to 0.76. The model with the poorest results is DT with a median of 0.69 and a high variance from 0.59 to 0.82 (Fig. 1).

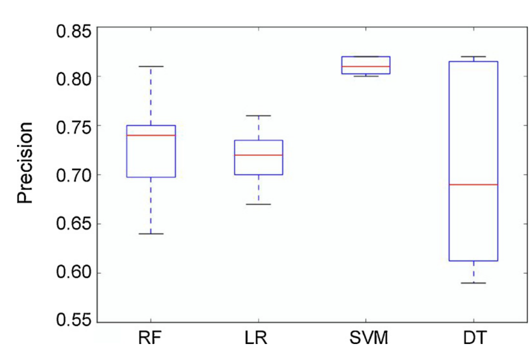

Fig. 1. ‘Precision’ from 10 steps of training and testing. RF = Random forest, LR = Logistic regression, SVM = Support vector machine, DT = Decision tree.

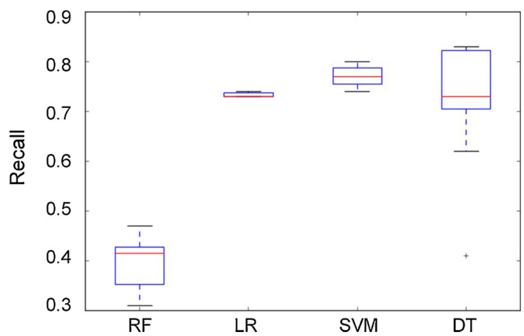

Fig. 2. 'Recall' from 10 steps of training and testing. RF = Random forest, LR = Logistic regression, SVM = Support vector machine, DT = Decision tree.

For the index 'recall', SVM has the highest median of all models with a value of 0.77 and a small variance from 0.74 to 0.80. LR and DT show similar results with a median of 0.74, but differ considerably in variance: LR 0.74–0.76; DT 0.65–0.84, with outliers of up to 0.42. The poorest results are represented by RF, with a median of 0.42, ranging from 0.31 to 0.47 (Fig. 2).

The results of the evaluation of the spatial simulation (Table 2, Fig. 3) were calculated according to Eqs. (3) and (4) (see chapter 2.4.2).

The following main results can be inferred from the Hasse diagram (Fig. 3):

1 The evaluation results in a clear structure.
2 DT is the model with the poorest results.
3 Combinations of two to three models yield the best results.
4 The combination of all four models performs worse than combinations of less models or even single models.

3.2. Secondary quality indices

LR and DT have the fastest algorithms (69 and 91 s, respectively), SVM has a medium (171 s) and RF the slowest runtime (518 s).

DT and RF are the only models that allow following and visualising the path of how the input variables were used for classification. LR and SVM are black boxes.

All models possess a lot of possibilities to fit them for the used dataset. We had to tune some of the default settings (e.g. number of estimators, depth of tree etc.) until the result of the training could not be improved anymore.

Table 2
'Exactness' (*E*, Eq. (4)) and 'selectivity' (*S*, Eq. (3)) of the used models and their combinations. RF = Random forest, LR = Logistic regression, SVM = Support vector machine, DT = Decision tree.

Model	Exactness	Selectivity
LR & RF	0.6303	0.6667
RF & SVM	0.6477	0.6348
LR & RF & SVM	0.6366	0.6398
LR & SVM	0.6085	0.6362
SVM	0.6270	0.6111
LR	0.5535	0.6455
RF	0.6122	0.5628
DT & LR & SVM	0.6064	0.5499
DT & RF & SVM	0.6415	0.5127
DT & LR & RF	0.6258	0.4018
DT & LR & RF & SVM	0.6178	0.3934
DT & SVM	0.6080	0.3956
DT & RF	0.6229	0.3879
DT & LR	0.5975	0.3926
DT	0.4858	0.2233

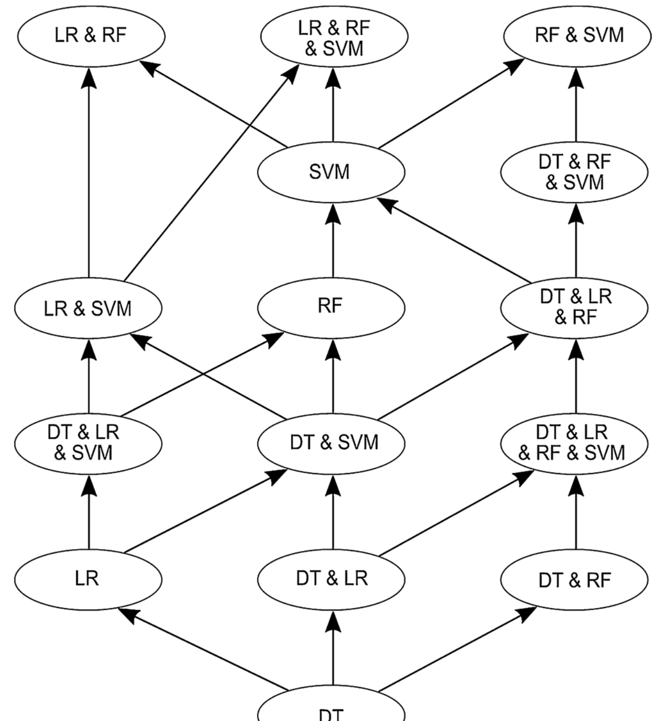

Fig. 3. Hasse diagram of 'selectivity' and 'exactness' (resulting from values of Table 2). Models or combinations of models on the same horizontal level are not comparable to each other. An arrow points to a better performing model or combination of models.

3.3. Spatial simulation

All selected models have in common, that they predict high probabilities of occurrence (0.7–0.9) for almost the whole area of the federal states of Baden-Wuerttemberg (BW) and North Rhine-Westphalia (NRW). The highest values are calculated by SVM, the lowest by DT. In other German federal states we find a high variation of the calculated prediction values. RF shows highly fragmented areas with values > 0.5 for Hesse (H), Rhineland-Palatinate (RP), Lower Saxony (LS), Thuringia (T) and Bavaria (BA). By contrast, SVM seems to calculate a more or less continuous area and very small additional spots in BA, but overall rather high prediction values of > 0.7.

LR looks similar to SVM, but predicts a much smaller distribution area and values for RP, but very high values of > 0.8 for the foothills of the Alps in southern BA. DT shows the biggest calculated area, with prediction values from 0.4 to 0.5, which includes LS, Schleswig-Holstein (SH), northern Saxony-Anhalt (SA), western parts of Brandenburg (BR), Mecklenburg-Western Pomerania (MP) and Berlin (B). Smaller areas with higher values are calculated by DT with the focus on the borders of BA and T, NRW, RP and H and some spots in BA (Fig. 4).

Most of the sites where individuals of *Ae. j. japonicus* were collected from 2011 to 2014 (BW and NRW) were correctly calculated by all models. Looking at more recently detected presence areas, e.g. from BA (first records from 2015; Zielke et al., 2016), the models show marked differences: The predicted probability of occurrence of *Ae. j. japonicus* calculated through DT and RF shows no or only weak probability values as compared to field collections from 2015 in the southern part of BA, when LR provides values of ~0.5–0.7 (Fig. 4).

The combination of models results in maps with areas of predicted occurrence that are more balanced than those from single models, because they are rid from outliers and much less fragmented. Frequently calculated high values are amplified, single weaker values are

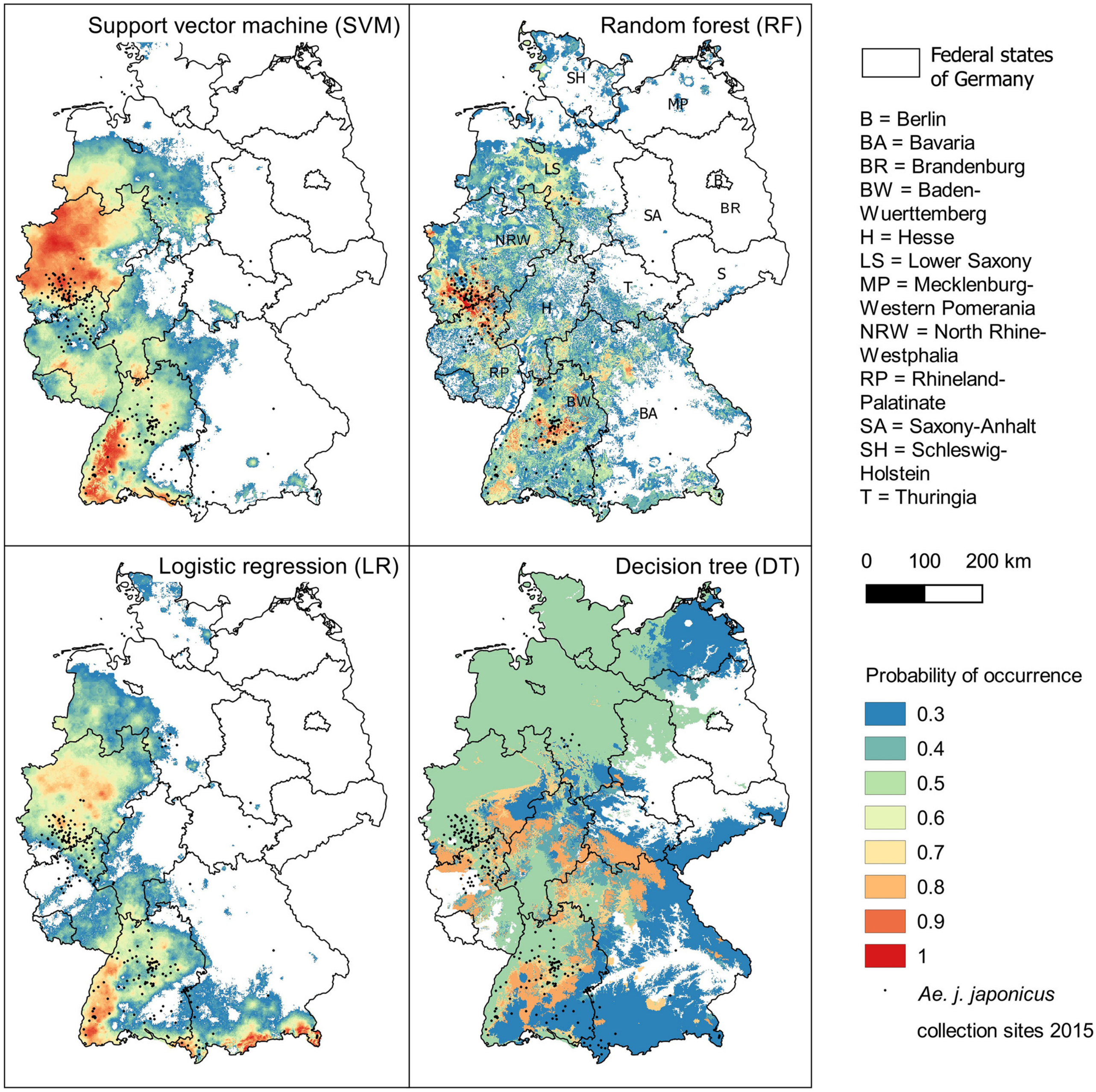

Fig. 4. Predicted probabilities of *Ae. j. japonicus* occurrence by the specific models, calculated with weather data from 2014 and mosquito training data from 2011 to 2014, as compared to field collection data from 2015.

diminished from the map. A negative effect can be caused by a single model with very poor results; it can bias the output from a combination of models with good evaluation results in such a way that it degrades the whole set (e.g. the addition of DT; Fig. 5).

When we look at the best results according to the Hasse diagram, which are the combinations of 'RF and LR and SVM', 'RF and SVM' or 'LR and RF', we see two main differences among many similarities: the combination 'RF and SVM' shows higher values for LS, but is much weaker for southern BA; by contrast, 'LR and RF' shows much higher values for southern BA (Fig. 5).

3.4. Comparison

The modelling outputs can be compared by summarising the different evaluation results. This is common in environmental modelling to get an overview about the quality of the models (e.g. Kampichler et al., 2010). In this case, each quality index contributes equally to the final result. Another option would be weighted ranking, which can be done independently if the modelling priorities are different (e.g. if false negatives should be absolutely avoided, the 'recall' index should have a higher value than the 'precision' index; or if there is a bigger dataset, the running time becomes more important).

According to the summarised results of the evaluation indices in Table 3, SVM can be ranked as the best model, due to its good results in the primary evaluation indices, and despite the facts that it is a black box model and rather slow. LR is faster than SVM, but is ranked second because it has weaker 'recall' and 'precision' results. RF follows third due to its slow runtime and a lower value in the index 'recall', probably

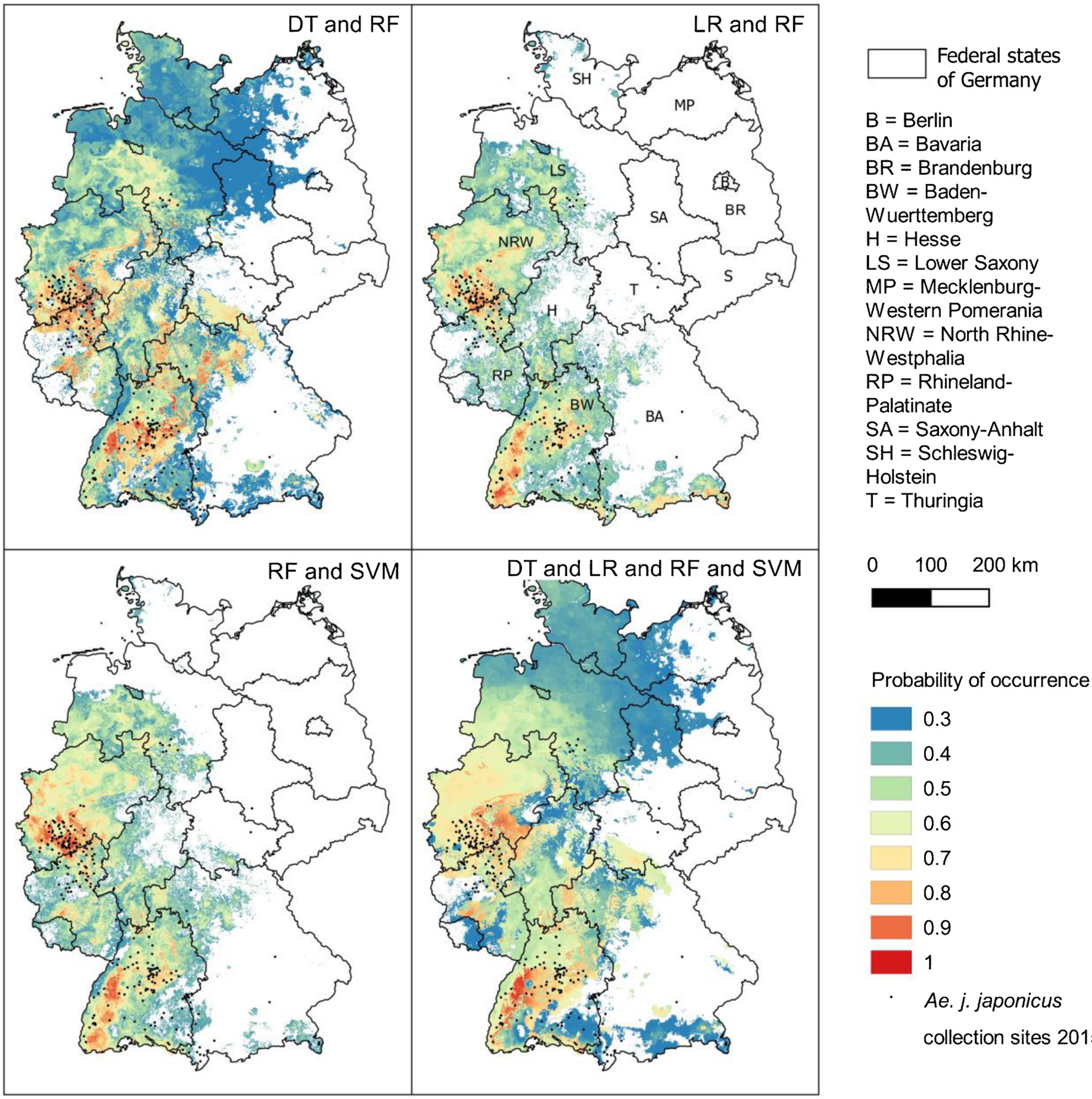

Fig. 5. Combinations of models, showing predicted probabilities of occurrence of *Ae. j. japonicus*, calculated with weather data from 2014 and mosquito training data from 2011 to 2014, as compared to field collection points from 2015.

owing to problems with the dataset or the training size (see chapter 4.1). DT is presented at the end because its variance makes it unreliable. However, the indices listed in Table 3 are incomparable. Therefore, the summarised results of the evaluation indices should be taken with caution. Instead for an accurate comparison of the models it is advised to prefer the results from the Hasse diagram (chapter 3.1).

4. Discussion

4.1. Performance

All models show good 'precision' index results, with the exception of DT which shows a rather high variance of this index, creating the

Table 3
Values of evaluation indices of single models, ranking from 1 (best) to 4 (worst). An asterisk (*) indicates high variance, resulting in poor values. The overall performance (result) is calculated as mean of the single indices.

Model	Recall	Value	Precision	Value	Time [s]	Value	Compre-hensibility	Value	Result
SVM	0.77	1	0.82	1	171	3	Black box	3	**2**
LR	0.75	2	0.72	3	69	1	Black box	3	**2.25**
RF	0.42	3	0.74	2	518	4	White box	1	**2.5**
DT	0.7*	4*	0.68*	4*	91	2	White box	1	**2.75**

impression that this model produces overfitted and unreliable output. An explanation could be the specific dataset, where locally clustered occurrence points can lead to an improper result. DT is prone to overfitting anyway, producing rather inaccurate and variable outputs with certain datasets compared to other models (James et al., 2013; Géron, 2017).

Regarding the 'recall' index, all models show good results, with the exception of RF. This could be the result of the randomised DTs (number of estimators = 10) in the training process. When the calculated maps are considered, the output of RF shows scattered, tessellated projected distribution areas, even in regions where a more contiguous area was expected. Such a mosaic-like pattern could be the reason for missing many of the testing points from 2015 and conveys the impression that the model is rather insensitive.

When we changed the number of estimators, we got the worse results for the classification performance the more estimators we included. This is a well-known problem in machine learning and is attributed to the bias-variance trade-off (Hastie et al., 2009). As RF reduces the variance of the single DTs, the bias of the model increases, resulting in underfitting and becoming overly conservative. This makes RF rather unreliable for helping detect unknown occurrence outside of the area of training data origin. On the other hand, it reduces the risk of calculating false positives. When we changed the training size from 1000 to 300 individuals, the 'recall' index became much better, but led to a rather liberal, unselective model, probably resulting from the insufficient number of training data and/or unbalanced input data (Subramanian, 2015). With a bigger dataset we assume that RF should yield a much better result, especially if some of the parameters for regularisation are adjusted (e.g. max_depth, min_samples_leaf, max_leaf_nodes).

For evaluating the modelling performance, the concept of partial ordering (Halfon, 1985), visualised through the HDT (Brüggemann et al., 1995), proves to be highly efficient. The approach of using the HDT to support the common evaluation method is new in entomological modelling studies and seems to be a promising tool for ranking model output.

Comparing the results from calculating the HDT (via the partially ordered set of 'selectivity' and 'exactness' indices), the combinations of 'LR and RF', 'SVM and RF' and 'SVM and LR and RF' produced the best results. Specific ensembles, containing two or three models, performed better than single models. The combination of all studied models was outperformed by most of the combinations with two or three models and also by some single models, e.g. SVM. Similar results, evaluated by other approaches, were recently obtained. For example a combination of up to 10 out of 27 models improved the training output (Martre et al., 2015), and subsets of five to seven models are outscoring the whole ensemble of 13 models (Solazzo and Galmarini, 2014).

Regarding the time effort for modelling, the rather simple structure of LR and DT may explain why these algorithms were the fastest and why RF with ten estimators was five to seven times slower. As the calculation in our study was already quite time-consuming, despite a rather small dataset, it is suggested to use LR or SVM for bigger datasets.

When the comprehensibility of the models is compared, we see that SVM and LR are black box models. In the white box models DT and RF, the importance of variables can be ranked (Breiman, 2001). Thus, in future modelling studies, the use of DT and RF can help to get a better understanding of the influence of specific variables.

Optimising the model by changing its setting does not only require understanding the model, but a well described documentation, which can be found for the scikit-package (Pedregosa et al., 2011). A certain setting can improve the accuracy of the model, but sometimes with negative side effects like an increase in false positive or false negative prediction. For example, we found the problem of overfitting in DT with exceedingly complex trees, which tend to learn incorrect patterns (in particular with the default setting: 'max_depth = None', resulting in a fully grown tree). Anyway, it is necessary to change the adjustments of each model before the application of the default settings, or else the output could be prone to serious overfitting (Rodda et al., 2011). All models showed good results after some tuning, also owing to a precise documentation (Pedregosa et al., 2011; Garreta and Moncecchi, 2013).

4.2. Further modelling steps

Weather variables alone cannot explain the occurrence of *Ae. j. japonicus*. It is necessary to include further aspects, such as land use, altitude and host population density (Liu et al., 2017) to get a more accurate picture of the possible distribution of *Ae. j. japonicus*. An approach to find such additional variables could be preliminary cluster computing, which has shown to be more promising than the selection by biological expertise (Wieland et al., 2017).

When simulating the spread of this species, rivers, roads and railways as pathways to distant locations, which may be climatically suitable, but not accessible by active migration of the mosquito, should also be analysed (Tannich, 2015; Holloway and Miller 2017). To solve the problem of adaptability of the species to the new environment, which produces uncertainties in specific variables, it is proposed to use a fuzzy modelling approach (Wieland and Mirschel, 2008; Costa et al., 2015).

5. Conclusion

By using eight weather variables and a dataset of different mosquito species, all four models calculate accurate predictions of the potential occurrence of *Ae. j. japonicus* in Germany, with some limitations for the rather unstable decision tree.

The evaluation of the spatial simulation with the introduced indices 'exactness' and 'selectivity' results in a partially ordered set. The application of the Hasse diagram technique visualises the partial set of 'selectivity' and 'exactness' indices of all models and combinations. Therefore, this technique appears to be a convenient instrument for the evaluation of machine learning methods.

Ensemble models compared to single models can give a more balanced and more diverse output, which is helpful to get information about previously unnoticed mosquito populations. Interestingly, the best results for the prediction of the species distribution are from combinations of two or three models. The combination of all available models is not advised, because there will always be some with weaker or instable output, which disqualify the output of the whole ensemble.

Acknowledgements

This paper was funded by the German Federal Ministry of Food and Agriculture (BMEL, grant no. 2819105315) and the Ministry of Science, Research and Culture of the Federal State of Brandenburg (MWFK), Germany. We are grateful to the German Weather Service for providing us with public weather data.

References

Bennett, N.D., Croke, B.F., Guariso, G., Guillaume, J.H., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., 2013. Characterising performance of environmental models. Environ. Model. Softw. 40, 1–20. https://doi.org/10.1016/j.envsoft.2012.09.011.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Breiman, L., Friedman, J., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth International Group, Belmont (CA) 368 pp.

Brüggemann, R., Steinberg, C., 1998. Partielle Ordnung. Umweltwissenschaften und Schadstoff-Forschung 10 (6), 376–378. https://doi.org/10.1007/BF03037684.

Brüggemann, R., Münzer, B., Halfon, E., 1994. An algebraic/graphical tool to compare ecosystems with respect to their pollution – the German river "Elbe" as an example – I: Hasse-diagrams. Chemosphere 28 (5), 863–872. https://doi.org/10.1016/0045-6535(94)90003-5.

Brüggemann, R., Schwaiger, J., Negele, R.D., 1995. Applying Hasse diagram technique for the evaluation of toxicological fish tests. Chemosphere 30 (9), 1767–1780. https://doi.org/10.1016/0045-6535(95)00061-C.

Buckland, S.T., Elston, D.A., 1993. Empirical models for the spatial distribution of wildlife. J. Appl. Ecol. 30 (3), 478–495. https://doi.org/10.2307/2404188.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297. https://doi.org/10.1007/BF00994018.

Costa, H., Ponte, N.B., Azevedo, E.B., Gil, A., 2015. Fuzzy set theory for predicting the potential distribution and cost-effective monitoring of invasive species. Ecol. Model. 316, 122–132. https://doi.org/10.1016/j.ecolmodel.2015.07.034.

Cox, D.R., 1958. The regression analysis of binary sequences. J. R. Stat. Soc. Ser. B (Methodol.) 20 (2), 215–242.

Cunze, S., Koch, L.K., Kochmann, J., Klimpel, S., 2016. *Aedes albopictus* and *Aedes japonicus* – two invasive mosquito species with different temperature niches in Europe. Parasit. Vectors 9 (1), 573. https://doi.org/10.1186/s13071-016-1853-2.

Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., 2014. Do we need hundreds of classifiers to solve real world classification problems? J. Mach. Learn. Res. 15 (1), 3133–3181.

Fischer, D., Thomas, S.M., 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 (1), 54–64. https://doi.org/10.1016/j.gloplacha.2011.05.008.

Fischer, D., Thomas, S.M., Neteler, M., Tjaden, N.B., Beierkuhnlein, C., 2014. Climatic suitability of *Aedes albopictus* in Europe referring to climate change projections: comparison of mechanistic and correlative niche modelling approaches. Eurosurveillance 19 (6). https://doi.org/10.2807/1560-7917.ES2014.19.6.20696.

Garreta, R., Moncecchi, G., 2013. Learning Scikit-learn: Machine Learning in Python. Packt Publishing Ltd., Birmingham 103 pp.

Géron, A., 2017. Hands-on Machine Learning with Scikit-learn & TensorFlow. O'Reilly Media, Sebastopol, USA, pp. 167–170.

Grenouillet, G., Buisson, L., Casajus, N., Lek, S., 2011. Ensemble modelling of species distribution: the effects of geographical and environmental ranges. Ecography 34 (1), 9–17. https://doi.org/10.1111/j.1600-0587.2010.06152.x.

Guisan, A., Zimmermann, N.E., 2000. Predictive habitat distribution models in ecology. Ecol. Model. 135 (2), 147–186. https://doi.org/10.1016/S0304-3800(00)00354-9.

Halfon, E., 1985. Is there a best model structure? III. Testing the goodness of fit. Ecol. Model. 27 (1–2), 15–23. https://doi.org/10.1016/0304-3800(85)90022-5.

Hasse, H., 1952. Über die Klassenzahl Abelscher Zahlkörper. Akademie-Verlag, Berlin 137 pp.

Hastie, T., Tibshirani, R., Friedman, J., 2009. Unsupervised learning. The Elements of Statistical Learning. Springer, New York, pp. 485–585.

Holloway, P., Miller, J.A., 2017. A quantitative synthesis of the movement concepts used within species distribution modelling. Ecol. Model. 356, 91–103. https://doi.org/10.1016/j.ecolmodel.2017.04.005.

Hosmer, D., Lemeshow, S., 2000. Applied Logistic Regression, 2nd edition. Wiley, New York 392 pp.

Huber, K., Pluskota, B., Jöst, A., Hoffmann, K., Becker, N., 2012. Status of the invasive species *Aedes japonicus japonicus* (Diptera: Culicidae) in southwest Germany in 2011. J. Vector Ecol. 37 (2), 462–465. https://doi.org/10.1111/j.1948-7134.2012.00252.x.

Huber, K., Jansen, S., Leggewie, M., Badusche, M., Schmidt-Chanasit, J., Becker, N., Tannich, E., Becker, S.C., 2014. *Aedes japonicus japonicus* (Diptera: Culicidae) from Germany have vector competence for Japan encephalitis virus but are refractory to infection with West Nile virus. Parasitol. Res. 113 (9), 3195–3199. https://doi.org/10.1007/s00436-014-3983-9.

James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. Tree-based methods. In: In: James, G., Witten, D., Hastie, T., Tibshirani, R. (Eds.), An Introduction to Statistical Learning. Springer Texts in Statistics, vol. 103 Springer, New York p. 315.

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* 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., Kuhlisch, C., Fröhlich, A., Scheuch, D.E., Walther, D., 2016. 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 (12). https://doi.org/10.1371/journal.pone.0167948.

Kampichler, C., Wieland, R., Calmé, S., Weissenberger, H., Arriaga-Weiss, S., 2010. Classification in conservation biology: a comparison of five machine-learning methods. Ecol. Inf. 5 (6), 441–450. https://doi.org/10.1016/j.ecoinf.2010.06.003.

Kraemer, M.U.G., Sinka, M.E., Duda, K.A., Mylne, A.Q.N., Shearer, F.M., Barker, C.M., Moore, C.G., Carvalho, R.G., Coelho, G.E., Van Bortel, W., Hendrickx, G., Schaffner, F., Elyazar, I.R.F., Teng, H.-J., Brady, O.J., Messina, J.P., Pigott, D.M., Scott, T.W., Smith, D.L., Wint, G.R.W., Golding, N., Hay, S.I., 2015. The global distribution of the arbovirus vectors *Aedes aegypti* and *Ae. albopictus*. eLife 4, e08347,. https://doi.org/10.7554/eLife.08347.

Liu, F., McShea, W.J., Li, D., 2017. Correlating habitat suitability with landscape connectivity: a case study of Sichuan golden monkey in China. Ecol. Model. 353, 37–46. https://doi.org/10.1016/j.ecolmodel.2016.09.004.

Lobo, J.M., Jiménez-Valverde, A., Real, R., 2008. AUC: a misleading measure of the performance of predictive distribution models. Glob. Ecol. Biogeogr. 17 (2), 145–151. https://doi.org/10.1111/j.1466-8238.2007.00358.x.

Martre, P., Wallach, D., Asseng, S., Ewert, F., Jones, J.W., Rötter, R.P., Boote, K.J., Ruane, A.C., Thorburn, P.J., Cammarano, D., Hatfield, J.L., Rosenzweig, C., Aggarwal, P.K., Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Brisson, N., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Grant, R.F., Heng, L., Hooker, J., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Kersebaum, K.C., Müller, C., Kumar, S.N., Nendel, C., O'leary, G., Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M.A., Shcherbak, I., Steduto, P., Stöckle, C.O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., White, J.W., Wolf, J., 2015. Multimodel ensembles of wheat growth: many models are better than one. Glob. Change Biol. 21 (2), 911–925. https://doi.org/10.1111/gcb.12768.

Melaun, C., Werblow, A., Cunze, S., Zotzmann, S., Koch, L.K., Mehlhorn, H., Dörge, D., 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 (3), 1051–1061. https://doi.org/10.1007/s00436-014-4274-1.

Olden, J.D., Lawler, J.J., Poff, N.L., 2008. Machine learning methods without tears: a primer for ecologists. Q. Rev. Biol. 83 (2), 171–193. https://doi.org/10.1086/587826.

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.

Pino-Mejías, R., Cubiles-de-la-Vega, M.D., Anaya-Romero, M., Pascual-Acosta, A., Jordán-López, A., Bellinfante-Crocci, N., 2010. Predicting the potential habitat of oaks with data mining models and the R system. Environ. Model. Softw. 25 (7), 826–836. https://doi.org/10.1016/j.envsoft.2010.01.004.

Podgorelec, V., Kokol, P., Stiglic, B., Rozman, I., 2002. Decision trees: an overview and their use in medicine. J. Med. Syst. 26 (5), 445–463. https://doi.org/10.1023/A:1016409317640.

Reggiani, M.G., Marchetti, F.E., 1975. On assessing model adequacy. IEEE Trans. Syst. Man Cybern. 5 (3), 322–330. https://doi.org/10.1109/TSMC.1975.5408407.

Rodda, G.H., Jarnevich, C.S., Reed, R.N., 2011. Challenges in identifying sites climatically matched to the native ranges of animal invaders. PLoS One 6 (2). https://doi.org/10.1371/journal.pone.0014670.

Sardelis, M.R., Dohm, D.J., Pagac, B., Andre, R.G., Turell, M.J., 2002a. Experimental transmission of eastern equine encephalitis virus by *Ochlerotatus j. japonicus* (Diptera: culicidae). J. Med. Entomol. 39 (3), 480–484. https://doi.org/10.1603/0022-2585-39.3.480.

Sardelis, M.R., Turell, M.J., Andre, R.G., 2002b. Laboratory transmission of La Crosse virus by*Ochlerotatus j. Japonicus* (Diptera: Culicidae). J. Med. Entomol. 39 (4), 635–639. https://doi.org/10.1603/0022-2585-39.4.635.

Sardelis, M.R., Turell, M.J., Andre, R.G., 2003. Experimental transmission of St. Louis encephalitis virus by *Ochlerotatus j. Japonicus*. J. Am. Mosq. Control Assoc. 19 (2), 159–162.

Schaffner, F., Vazeille, M., Kaufmann, C., Failloux, A.B., Mathis, A., 2011. Vector competence of *Aedes japonicus* for chikungunya and dengue viruses. Eur. Mosq. Bull. 29, 141–142.

Solazzo, E., Galmarini, S., 2014. Multi-model ensembles: how many models do we need? In: Steyn, D., Mathur, R. (Eds.), Air Pollution Modeling and its Application XXIII. Springer International Publishing, pp. 505–510. https://doi.org/10.1007/978-3-319-04379-1_83.

Subramanian, G., 2015. Python Data Science Cookbook. Packt Publishing, Birmingham 438 pp.

Takashima, I., Rosen, L., 1989. Horizontal and vertical transmission of Japanese encephalitis virus by *Aedes japonicus* (Diptera: culicidae). J. Med. Entomol. 26 (5), 454–458. https://doi.org/10.1093/jmedent/26.5.454.

Tannich, E., 2015. In: Umweltbundesamt (Ed.), Auswirkungen des Klimawandels auf die Verbreitung krankheitsübertragender Tiere: Importwege und Etablierung invasiver Mücken in Deutschland. Umwelt & Gesundheit, 09/2015, Dessau-Roslau 36 pp.

Tarca, A.L., Carey, V.J., Chen, X.W., Romero, R., Drăghici, S., 2007. Machine learning and its applications to biology. PLoS Comput. Biol. 3 (6). https://doi.org/10.1371/journal.pcbi.0030116.

Thomson, M.C., Doblas-Reyes, F.J., Mason, S.J., Hagedorn, R., Connor, S.J., Phindela, T., Morse, A.P., Palmer, T.N., 2006. Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. Nature 439 (7076), 576–579. https://doi.org/10.1038/nature04503.

Tsonkova, P., Böhm, C., Quinkenstein, A., Freese, D., 2015. Application of partial order ranking to identify enhancement potentials for the provision of selected ecosystem services by different land use strategies. Agric. Syst. 135, 112–121. https://doi.org/10.1016/j.agsy.2015.01.002.

Vezza, P., Muñoz-Mas, R., Martinez-Capel, F., Mouton, A., 2015. Random forests to evaluate biotic interactions in fish distribution models. Environ. Model. Softw. 67, 173–183. https://doi.org/10.1016/j.envsoft.2015.01.005.

Voigt, K., Brüggemann, R., Pudenz, S., 2006. A multi-criteria evaluation of environmental databases using the Hasse Diagram Technique (ProRank) software. Environ. Model. Softw. 21 (11), 1587–1597. https://doi.org/10.1016/j.envsoft.2006.05.005.

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). https://doi.org/10.1093/jme/tjx166.

Werner, D., Kampen, H., 2013. The further spread of *Aedes japonicus japonicus* (Diptera, Culicidae) towards northern Germany. Parasitol. Res. 112 (10), 3665–3668. https://doi.org/10.1007/s00436-013-3564-3.

Werner, D., Hecker, S., Luckas, M., Kampen, H., 2014. The citizen science project 'Mückenatlas' supports mosquito (Diptera, Culicidae) monitoring in Germany. In: Müller, G., Pospischil, R., Robinson, W.H. (Eds.), Proceedings of the 8th International Conference on Urban Pests. 20–23 July 2014, Zurich, Switzerland. pp. 119–124.

Wieland, R., Mirschel, W., 2008. Adaptive fuzzy modeling versus artificial neural networks. Environ. Model. Softw. 23 (2), 215–224. https://doi.org/10.1016/j.envsoft.2007.06.004.

Wieland, R., Voss, M., Holtmann, X., Mirschel, W., Ajibefun, I., 2006. Spatial analysis and modeling tool (SAMT): 1. Structure and possibilities. Ecol. Inform. 1 (1), 67–76.

https://doi.org/10.1016/j.ecoinf.2005.10.005.
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. https://doi.org/10.1016/j.ecoinf.2015.08.002.
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.
Zielke, D.E., 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, 163. https://doi.org/10.1186/s13071-016-1447-z.