Additive Compendium Map of Outbreak Risk Determinants of West Nile Virus in Europe at NUT3

Alberto Alexander Gayle[1]

1. Department of Public Health and Clinical Medicine, Section of Sustainable Health, Umeå University, SE-90187 Umeå, Sweden

corresponding author(s): Alberto Alexander Gayle (albertoagayle@gmail.com)

Abstract

The spread of West Nile virus among humans is dependent on a wide range of determinants. However, predictive efforts are confounded by complex interactions and time-varying associations. SHAP (SHaply Additive Explanation) is a novel AI-driven solution with potential to overcome this. SHAP takes a high-performance XGBoost model and deductively imputes feature effects and predictive structure on a case-wise basis. The derived effect matrix is dimensionally identical to the original data set and can be summarized to produce estimates and analyses particular to any given subset of cases or features. Here, SHAP was applied to a database consisting of high-resolution data from various domains – climate, environment, economic, sociodemographic, vector and host distrubtion – to derive a feature effect matrix predictive of local (NUTS3) WNV outbreak risk in Europe. The extraordinary WNV outbreak of 2018 was used for out-of-sample validation. Four prospective models confirmed overall accuracy and stability of the resulting effect matrix. Further diagnostics confirmed the statistical robustness of this SHAP-derived data set with respect to various analyses, both individually and grouped.

Background & Summary

West Nile Virus (WNV) is a mosquito-borne virus with deadly potential[1]. The virus has since rapidly spread in recent years and now enjoys the widest geographic distribution of any similar mosquito-borne viruse[2]. 2018 alone saw a 7.2 fold increase in the number of cases reported in Europe along with a markedly expanded geographic range (see *Figure 1*). Many factors have been associated with this spread, but much is still unknown[3]. The endemic potential of WNV is not to be underestimated: seropositivity is reportedly as high as 90% in the sub-Saharan regions where WNV has long been endemic[4]. Reports of human-to-human transmission routes – via blood transfusion[5], organ transplants[6], aerosol[7,8], birthing and breast feeding[9], and possibly even sexual contact[10]– have added to the concern[11,12]. Response options are limited to symptomatic treatment and no human vaccine is available[10]. And the continued emergence of multiple lineages[13–16], each with distinct transmission and syndromatic profiles[17], adds to the uncertainties. Predicting annual emergence and spread of WNV in Europe has therefore become a priority.

The spread of West Nile virus is dependent on a wide range of determinants, including climatic features, environment, and sociodemographic factors[18]. However, predictive modeling efforts have often been confounded by complex interactions and time-varying associations[19]. Early warning models can therefore often be difficult to implement in practice[20]. Various analytical modalities have been used for this purpose, each with different pros and cons. Most have met with varying degrees of success, but scope- or context-dependent inconsistency of effect remains an outstanding issue[21]. Classification tree models are a promising entry in this field[22]. However, some controversy has been merited. Most notably, questions persist as to the long-term predictive accuracy of these models. This is due to the "black box" nature of these models whereby the mechanistic structure being assumed by these models are not open to scrutiny: unlike traditional models, no parameter estimates are provided[22]. This makes such models particularly unsuited for the early warning case, or intervention strategy, as 1) direct effect of specific predictors on the predictive outcomes cannot be determined, and 2) there is no way to control or investigate the reliability of the deterministic structure driving predictions.

SHAP (SHaply Additive Explanations)[23] is a novel AI-driven solution with potential to overcome overcome the problems associated with classification tree models. SHAP takes a high-performance XGBoost model[24] and deductively imputes feature effects and predictive structure on a case-wise basis[25]. The effect matrix derived from the resulting SHAP meta-model is dimensionally identical to the original data set and can be summarized to produce effect estimates and analyses particular to any given subset of cases or features. The underlying predictive engine, XGBoost, is notably robust against multicollinearity[26], automatically facilitates data curation[27], and outperforms even the most state-of-the-art prediction methods for most types of structured data[28]. SHAP has been mathematically proven to be optimal with respect to estimating feature effect contribution[29]. Provided sufficiently high predictive performance in the original XGBoost model, the determinant substructure revealed by SHAP can therefore be assumed to enjoy a high-degree of empirical validity – more so provided a sufficiently distinct validation frame.

Herein, we report the application of the XGBoost-SHAP pipeline to generate to a large, high-resolution data set consisting of feature effect variables predicitive of WNV from across various domains – climate, environment, economic, sociodemographic, vector distribution, and host presence. For validation, the extraordinary WNV outbreak of 2018 in Europe was selected as the predictive objective due to the extreme degree to which it differed from prior years. The overall accuracy and stability of SHAP-derived feature effect estimates is

demonstrated by comparing the output of various model constellations. Predictive quality of the SHAP-derived feature effect matrix is furthermore evaluated by comparing it to the original, nominal data with respect to traditional multivariate analyses, such as logistic regression. The scope of this resulting, validated dataset covers the training period, 2010-2016, and the 2018 validation year. For completeness, the 2017 hold-out year has also been included in the published data set.

Methods

Original data collection

A literature review was conducted. The purpose of this review was to identify potential covariates of West Nile virus (WNV) emergence. PubMed and Google Scholar were searched. All research published between 1999 and 2019 was identified that included "West Nile" in either the title or abstract. This set of research publications was furthermore filtered. Only those titles or abstracts that referred to geospatial risk associations were included. Search terms for this secondary filter included: "risk", "model", "predict", "estimat", "associat", "determinant", "factor" and "review". This list was then manually screened to exclude articles that presented neither primary research nor references to the results of such. The determinants presented were then extracted and compiled. Time-lag and other interaction effects were noted separately and in addition to the primary determinants. Determinant specifications with respect to aggregation, normalization, and other processing were also noted. This list was then stratified into eight categories based on class and source of data. Spatial and temporal covariates were included among the eight feature classes.

The selected feature classes included the following broad categories: "climate", "environmental", "economic", "sociodemographic", "hosts", and "vectors". Public data sources for each feature class were identified and selected based on the availability of data that aligned with determinant specifications. The literature on WNV consistently describes it as "hard to predict". And this is reflected in broad uncertainties surrounding the precise mechanistic associations surrounding even the most basic, and generally accepted, determinants. In many cases, time-lagged climatic features are included in models to serve as proxy for specific features relevant to vector breeding or disease transmission. For example, what does it mean to associate rain that occurred 4 weeks prior to a reported case today? What is the actual biotic or environmental feature that is driving increased risk today: a) precipitation accumulation, b) soil moisture, c) standing water, or d) increased surface reflectivity indicative of water presence (MNDWI)? In order to avoid such potentially confounding effects, we expanded our inclusion criteria to include all climatic or environmental indicators available from a given selected source – this included some features not present in the literature.

All data was downloaded and processed using the R platform. Spatial and temporal covariates were included: one for each of the district-level regions in Europe (NUT3) as well as years 2010-2016 and 2018. Geometries for each of the NUTS3 regions were obtained using the "Eurostats" package in R. Public sources were sought out for each of the remaining feature classes. See *Supplementary Table 1* for a complete summary of all included features, sources, processing specifications, and summary statistics. The "earthEngineGrabR"[30] package in R was used to obtain environmental and climate data from Google Earth Engine. Climatic trend data was obtained directly, in preprocessed form from the International Research Institute for Climate and Society (IRI) at Columbia University (New York, USA)[31]. The "Bioclimatic" feature set is a well-cited set of time-varying geoclimatic metrics applied in the

area of ecological niche modeling[32]. They are derived from standard climatic variables such as precipitation and temperature, over time, for a given region. The "dismo" package[33] in R was used to calculate these. In addition, indices for intradecal climatic systems such as the "El Nino South Oscillation" (ENSO) and the "North Atlantic Oscillation" (NAO) were obtained using the "rSOI"[34] package in R. To account for time-lag effects, all climatic variables were aggregated into four quarters as done in similarly scoped studies[35]. This excluded the Bioclimatic variables which are defined annually.

Environmental and host data was obtained directed from the European Environment Agency (EEA)[36]. This included all Corine Land Cover (CLC) and Natura 2000 Bioregions data[37]. Geospatial species distribution data was also obtained from the EEA, by merging the Natura 2000 Habitats and Species tables. This included all species, to allow for potentially unreported hosts. Economic and sociodemographic data was obtained using the "Eurostat" package[38]. The European Center for Disease Control (ECDC) provided mosquito distribution data; this included all reported vectors in the region. West Nile virus case data was also obtained from the ECDC.

WNV case data was quality-controlled to exclude cases likely to be subject to regional variation in reporting or other biases. Excluded were 1) asymptomatic cases, 2) cases with no syndromatic description, and 3) cases acquired outside the reporting district. This is consistent with previous work, in which all but the most severe neuroinvasive manifestations (WNND) were excluded[39].

This final data set was then divided into a training set (2010-2016) and test/validation set (2018). The statistical effect of large-scale climatic systems and associated downstream covariation may potentially overlap year-to-year. To mitigate, 2017 was omitted from training.

Synthetic data generation

The final data set was thus used to construct an explanatory model from which the SHAP effect matrix would be extracted. This included the full year's data, including climatic data from periods overlapping and extending beyond the outbreak season (i.e., January - December). This model was therefore not suited for same year prospective early warning, but would serve instead to retrospectively examine the factors driver the 2018 outbreak.

The official implementation for XGBoost in R[24] was used. Cross-validation was used to set the initial iteration limits. A training data set was generated for each using data from years 2010-2016, according to the specification of each model. For cross validation, a hold out subset is specified and then used to evaluate the performance of a model trained with the remaining data. Here, given the cyclical nature of our predictive task, the hold out set was specified to be a single year. For each model, the training set was therefore split according to year and a model was trained using six of the seven years. The seventh, hold out year was then used to test model performance. This process was repeated seven times, using each of the included years once as the hold out set. This process generates several benchmarks, one of which is an optimal number of iterations. For classification tree models, iteration limitations are commonly set so as to avoid over-fitting. A second model was trained using the entire seven years of training data, with the iteration limit set based on the limit assessed by the cross-validation run. An additional criteria was set to stop model training after 10 iterations with no reduction in log-loss error. Once this second training round was complete, covariates determined to be insignificant were dropped and the model was rerun. This mode of feature selection is consistent with a common use case for XGBoost[27] and it is

furthermore recommended for this class of model[40]. In our case, covariates found to be insignificant with respect to three importance criteria – marginal contribution to predictive output ("gain"), internal predictive prevalence ("frequency"), and case-wise predictive relevance ("cover") – were removed.

In a departure from established machine learning norms[41,42], model hyperparameters were not subject to automated optimization procedures. Models such as XGBoost are notoriously sensitive to hyperparameter settings[43] – a problem we wished to avoid outright. Inspired by the work of Blagus and Lusa (2017)[44] and Brieman (2001)[40], hyperparameters were therefore set to ensure internal experimental robustness: 1) 95% was set as the internal discriminatory threshold, 2) each submodel was limited to half of the total feature space, and 3) maximum tree size was set to twice the number of included feature classes (i.e., 16). Consistent with guidance put forth in the seminal paper by Elith J. et al (2008), step size was reduced to 0.03 (10% of the default) to ensure possible feature interactions were exhaustively assessed.

The "SHapley Additive exPlanations" (SHAP) engine is an AI-driven deductive framework for imputing local feature effects[23]. The underlying algorithm is derived from a military intelligence application based on cooperative game theory. This algorithm determines individual contributions for group outcomes and has been mathematically proven to be optimal with respect to this task[25]. The SHAP implementation is applied to an XGBoost model to determine the degree to which individual features influence the computed event probability for each case. The output of this process is a meta-model, consisting of an effect matrix dimensionally identical to the original, nominal value data (matrix): one complete set of feature effect estimations for each case. Each term reflects the local change in log relative risk associated with a given feature with respect to the global, baseline probability of event. This global baseline risk is computed along with the effect matrix and is the same for all cases. This transformation effectively recenters the predictive output at zero, implying a predictive threshold equivalent to the computed baseline probability. Values greater than zero indicate increased relative risk of event and *vice versa*. This transformation provides for several useful analytical modalities that are presented later and in the main text. The inputs for this process are the computed XGBoost model and the original, nominal value data (set).

The SHAP engine was therefore applied to this model to generate a SHAP meta-model for the 2018 validation period. The choice of classification threshold directly affects predictive performance and several methods exist to optimize[45]. However, as with manual or automated tuning of hyperparameters in XGBoost, the effect of such methods on out-of-sample performance is often unpredictable[46]. Out-of-sample performance was therefore evaluated at the threshold inherent to the SHAP transformation – zero. In-sample performance was likewise estimated using the 2010-2016 training set. For each, the following metrics were assessed: 1) AUC, 2) sensitivity, 3) specificity, and 4) balanced accuracy. AUC is the probability that a randomly selected positive case will be assigned a higher probability than a randomly selected negative case. It indicates how well the model discriminates between positive and negative classes irrespective of classification threshold. The remaining three metrics are threshold-dependent and vary accordingly. Sensitivity, also referred to as true positive rate, measures the probability of detecting event occurrence. Specificity, also referred to true negative rate, measures the probability of correctly identifying non-occurrence. Balanced accuracy is the average of the prior two metrics; it measures the probability of detection assuming negative and positives classes are equally represented in the data set. In the present study, positive classes represent only 0.05% of the total sample. A model that optimizes balanced accuracy irrespective of threshold is therefore preferable. The SHAP-implied classification threshold was also applied to the original XGBoost model, on a diagnostic basis, to confirm predictive equivalence.

For validation, logistic regression, linear regression and confirmatory path analysis were conducted for the 2018 validation period. Features were ranked according to mean absolute SHAP effect. Features were then added in step-wise fashion to a logistic regression model according to the SHAP-derived ranking, for both the original nominal value matrix and SHAP-derived effect matrix. The predictive objective was binary presence of WNV. For each step, pseudo R-squared was computed and compared, for p=1 through p=50 features. Features were subsequently grouped according to feature class and similarly analyzed figure step-wise logistic regression, for p=1 through p=30 features, for both the data sets. Each feature class set was subsequently aggregated and added to a linear regression model, with actual case count as the predictive objective. Goodness of fit was obtained along with parameter estimates for each aggregate. Confirmatory path analysis was subsequently applied to confirm empirical validity. Predictive objectives included were 1) actual case count and 2) outbreak risk level (log relative risk) as determined by SHAP.

The "XGBoost"[47] package in R was used for generation of the initial XGBoost models as well as the SHAP meta-models. Linear and logistic regression were conducted using the built-in R function "lm". Confirmatory path analysis was conducted and visualized using the "piecewiseSEM"[48] package in R. The "ggplot2"[49] package in R was used for visual summarization and the "tmap"[50] package in R was used for geospatial mapping. Geospatial polygons were obtained from the Statistical Office of the European Commission using the R package, "eurostat"[38]. All computations as well as as necessary data processing were conducted in R, using the "data.table"[51] package.

Limitations

Several *a priori* assumptions and modeling choices were required. Such choices naturally influence the final model and output in various ways, not limited to overall predictive performance and imputed associations with respect to individual features. Most notably, the final feature set size for any given model is in no way deterministic and could very well be a function of model hyperparameters, namely "tree-depth" and "eta". These two hyperparameters effectively control the degree to which the model explores causal alternatives and complexity of causation, and it is plausible to presume some effect. Furthermore, the data set used to derive this model was in no way comprehensive or statistically refined and outputs are sure to change given additional fine-tuning in this regard. That having been said, exploring the performance of these methods with data of uncertain or uneven quality was an implicit goal of this work.

Data Records

The global habitat map for the year 2018 (Fig. 21) is made available online via Github (https://github.com/aruberuku/EU2018WNV/) with an interactive version soon to follow. As part of this manuscript, the map for each of the six feature classes – climate, environment, economic, sociodemographic, vectors, and hosts – is included on the github page along with geospatial data at NUT3 resolution. The code to recreate the map is also available. Asset data used are publicly readable and directly available from the original sources (see methods). Users are advised to check the data repository for newer versions of both code and map, as we consider this product will be improved in the future pending better data availability. Further refinements to the methodology or feature class designations may also warrant further updates. A complete data compendium is provided as a supplement.

Technical Validation

Out-of-sample performance for the underlying XGBoost model was 96%, which far exceeds that of any reported, similarly-scoped model. Sensitivity, specificity, and balanced accuracy were 0.89, 0.96, and 0.92 respectively at the SHAP-implied classification threshold of 0.0062. Without further validation, the demonstrated degree of out-of-sample performance would imply an underlying determinant structure with a high degree of empirical validity. The following technical validation and discussion further qualifies this finding and demonstrates the degree to which the SHAP-derived effect matrix statistically and qualitatively out-performs the original, nominal value data.

Summary of SHAP specification and estimation process. Many methods exists for deriving statistical effect estimates for multivariate models and none are considered ideal. Introduced in 1951, the Shapely value estimation heuristic was designed to deliver optimal effect estimates – a quality that has since been mathematically demonstrated to be true. Shapely values represent that "expect marginal contribution of [a given feature] to any random coalition" and have been mathematically proven to be optimal with respect to this task []. Shapely values are heuristically determined according to the following axioms in Table 1:

Property	Specification	Implication
Efficiency	The sum of the Shapley values of all agents equals the value of the grand coalition, so that all the gain is distributed among the agents	For each local case, the overall predictive outcome is an additive function of the Shapely effect estimates; the system is therefore invariant to permutation
Symmetry	If two cases are functionally identical then the feature-wise distribution of Shapely values will be equal as well	Shapely effect estimates and associated variability reflects functional equivalency on a feature-wise basis, where applicable
Additivity	The aggregate or interactive effect of two features is equal to the sum of the individual Shapely values	The effect of any arbitrary aggregation of features is independently valid and meaningful, as far as the data allows
Null player	If the predictive outcome does not change on addition or removal of a feature, then Shapely value is zero	Shapely effect estimates are based on observed marginal contribution and are not subject to estimation artefacts.

The result of applying this heuristic are local feature effect predictions that are independent individually and on the aggregate. The efficiency and additivity properties therefore allow for meaningful decomposition of the SHAP effect matrix on a case- and feature-wise basis. The symmetry and null player properties furthermore imply that any statistical associations between features exposed by multivariate analyses will be meaningful with respect to any empirically observable relationships between said features and the predictive outcome.

SHAP effect matrix is statistically robust and empirically sound. As implemented in SHAP, local effects are deductively imputed based on the marginal contribution determined via reanalysis of the decision tree ensemble from which the original XGBoost predictions were derived. In the present model, each XGBoost prediction is backed by thousands of effect estimations across all statistically relevant combinations of features. Each local effect estimation is therefore supported by a broad range of experimental configurations, with feature value combinations and predictive output that is unique for each local case. One practical result is that the imputed effects for a given feature will vary independently, on a

case-wise basis, even if the original input feature values were the same. Lower resolution data will therefore be projected into the higher resolution predictive space, with an increase in empirical precision. Imputed SHAP effects are therefore both a) highly correlated with the original, nominal value data (Figure 1a) and b) overall better predictors than the nominal value data—even when estimated on an out-of-sample basis (Figure 1b).

[Figure 1a,b]
[Legend F1a,b]

Recall that all feature-wise constellations are equally and independently valid with respect to the predictive outcome. In the present case, six distinct feature classes were included as predictors and logistic regression revealed each set to independently meaningful with respect to outbreak risk prediction (Figure 2a), with the SHAP effect estimates consistently – and in some cases, substantially – outperforming the original, nominal value data, particularly at lower feature set sizes. And geospatial visualization provide qualitative confirmation: each domain-specific map reflects independently meaningful, domain-specific knowledge relevant to local outbreak risk (Figure 2b). Regression analysis furthermore confirmed both validity and statistical independence of domain-specific aggregations of the SHAP effect estimates even when applied to alternative outbreak criteria – in this case, total case count per region (Table 2). This is a considerably more sensitive predictive objective and the high-performance demonstrated speaks to the empirical robustness of the SHAP estimates. Empirical validity of the SHAP-derived outbreak risk map and implications with respect to local and cross-scale determinant relationships are treated in two separate, forthcoming analyses (preprints available). Others are welcome to explore and evaluate this data further.

[Figure 2a,b]
[Legend F2a,b]

[Table 2]
[Legend T2]

Usage Notes

The SHAP effect matrix is dimensionally identical to the original, nominal value data and is provided in tabular form. As demonstrated, any analytical steps or methods suitable for the original data would be equally valid with the SHAP effect matrix. Further analytical application are presented in a forthcoming manuscript, available as a pre-print online. Methods applied include: means testing (student's t-test) with respect to drivers of positive indication and 2018 outbreak risk, standardized effect estimation, and clustering analysis.

Others are welcome explore and evaluate this data further, with any applicable methods. Some suggestions might include regression to estimate SHAP-derived effects for individual vectors as a function of the nominal values of various climatic features. Another might be analyses to determine interaction effects between individual host, vectors, and classes of environment. Country and region-specific analyses are also possible, along with comparative analyses based on common feature ranges such as population density or proximity to water.

In additional, external data could also be married to this data set. In all cases, we recommend the R platform for ease of combined geospatial and statistical analysis. However, any other standard statistical software package will do – at least on the analytical side.

Code Availability

All programming code necessary to reproduce the map is supplied together with the data (see Data records) and on https://github.com/aruberuku/EU2018WNV/.

Acknowledgements

Funding was received from the European Center for Disease Control and Prevention (contract no ECDC.9504). We would like to thank the European Center for Disease Control for kindly providing funding, case and vector data, and support. And special thanks to the team at Exploratory.io who provided technical guidance and support.

Author contributions

The authors worked collaboratively and contributed equally to the article.

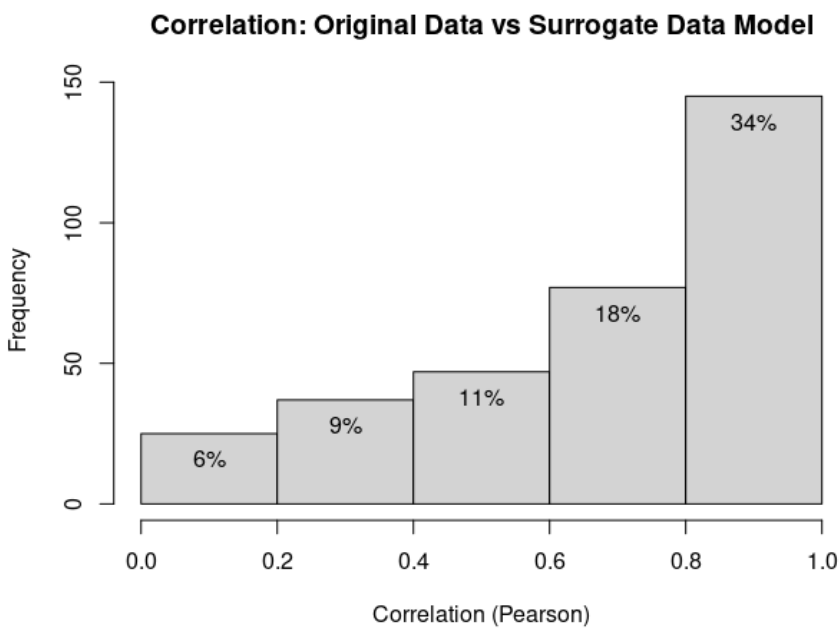
Competing interests

The authors declare no competing interests.

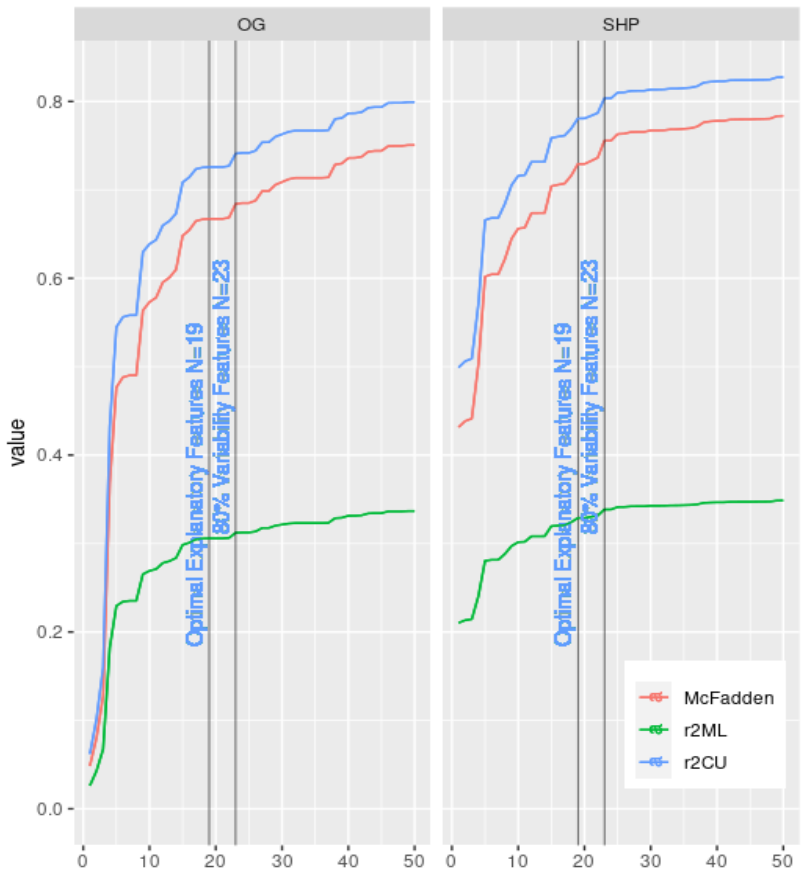
Figures

[Figure 1a,b]

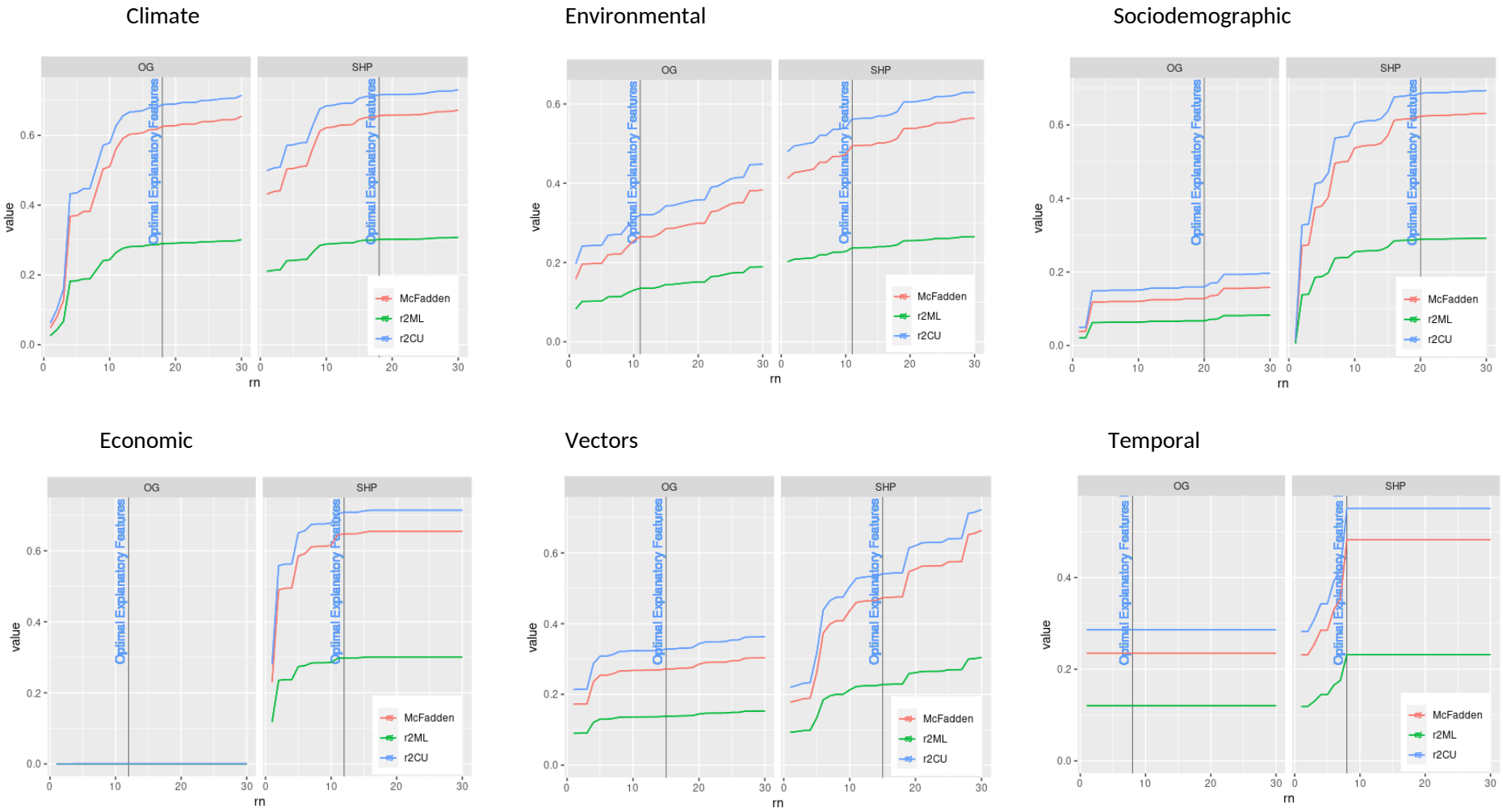
a)



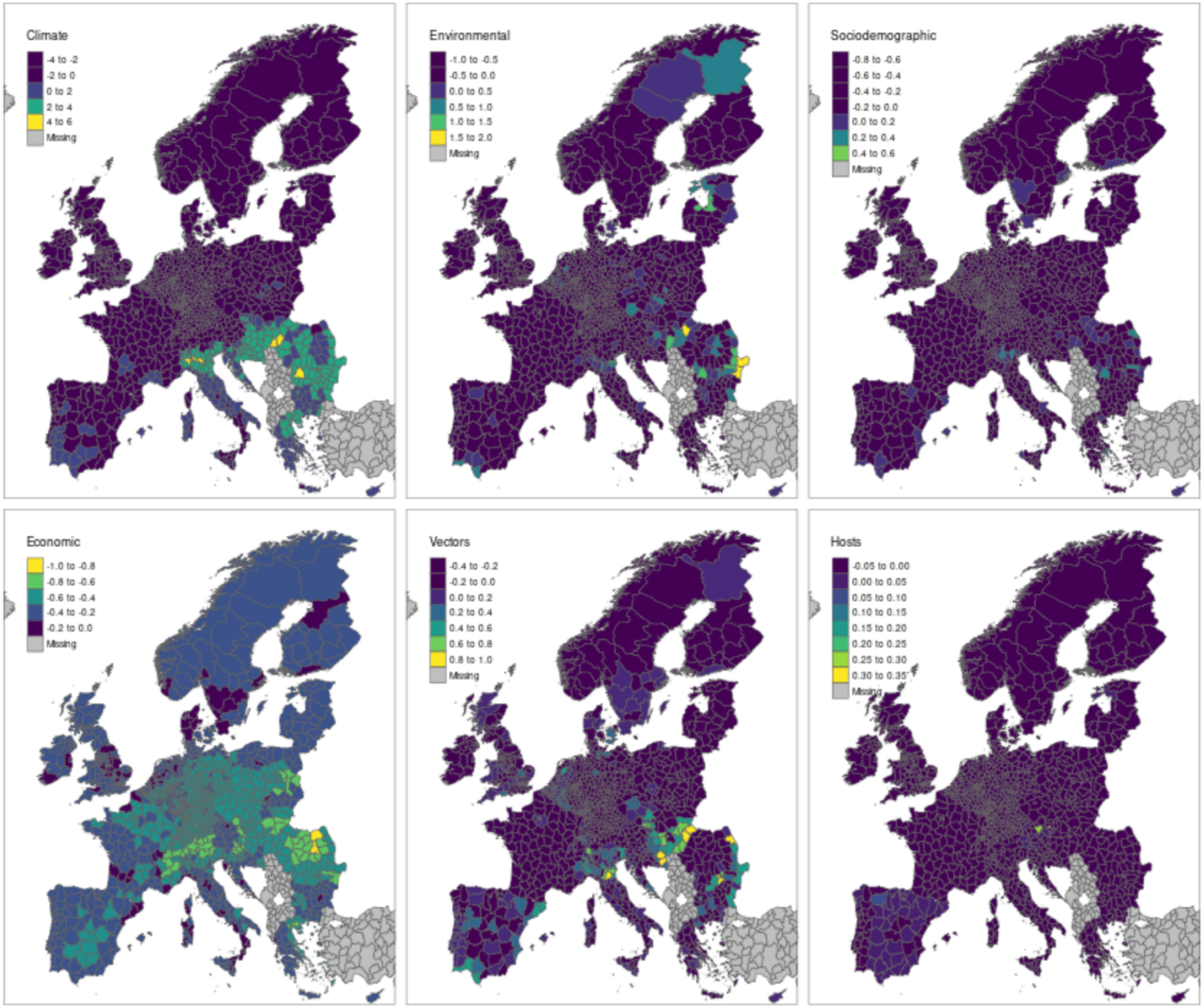
b)



[Figure 2a]


Climate
Environmental
Sociodemographic
Economic
Vectors
Temporal
OG
SHP
Optimal Explanatory Features
value
rn
McFadden
r2ML
r2CU

[Figure 2b]


Climate
-4 to -2
-2 to 0
0 to 2
2 to 4
4 to 6
Missing
Environmental
-1.0 to -0.5
-0.5 to 0.0
0.0 to 0.5
0.5 to 1.0
1.0 to 1.5
1.5 to 2.0
Missing
Sociodemographic
-0.8 to -0.6
-0.6 to -0.4
-0.4 to -0.2
-0.2 to 0.0
0.0 to 0.2
0.2 to 0.4
0.4 to 0.6
Missing
Economic
-1.0 to -0.8
-0.8 to -0.6
-0.6 to -0.4
-0.4 to -0.2
-0.2 to 0.0
Missing
Vectors
-0.4 to -0.2
-0.2 to 0.0
0.0 to 0.2
0.2 to 0.4
0.4 to 0.6
0.6 to 0.8
0.8 to 1.0
Missing
Hosts
-0.05 to 0.00
0.00 to 0.05
0.05 to 0.10
0.10 to 0.15
0.15 to 0.20
0.20 to 0.25
0.25 to 0.30
0.30 to 0.35
Missing

[Table 2]

	Estimate	Std. Error	t-value	Pr(>|t|)	
(Intercept)	3.51	0.45	7.75	0.00	***
Climate	0.80	0.07	12.16	0.00	***
Economic	-0.37	0.96	-0.39	0.70	
Environmental	0.84	0.31	2.72	0.01	**
Sociodemographic	3.40	0.89	3.81	0.00	***
Vectors	6.51	0.77	8.47	0.00	***
Hosts	-17.11	9.69	-1.77	0.08	.
R^2:	*0.230*				
Adj. R^2:	*0.228*				

Figure Legends

[*Legend F1a,b]*

Surrogate data model aligns with but outperforms original data. The surrogate data model is heuristically designed to reproduce the output of the original (XGB) model while estimating the individual, localized effect of each feature. The surrogate data model is therefore dimensionally identical to the original data set but reflects deductively approximated effect estimates. We confirmed a high-degree of statistical correspondence between the two (*Figure 1a*). Stepwise logistic regression (*Figure 1b*) furthermore confirmed that the variability encoded within the surrogate data model aligned better with the original case data, for any given feature set size. This observed difference in predictive performance notably increases as model dimensionality is reduced. This finding is consistent across under various scenarios of model decomposition – e.g, by feature class.

[*Legend F2a]*

Feature effects predicted by SHAP offer superior statistical robustness. A supplementary analysis was conducted in order to diagnostically assess the properties of the parameters generated by the Shapley transformation. Step-wise logistic regression series were generated for feature class; AUC performance is shown here plotted as a function of the number of included features; hosts showed anomalous results due to limited cases for each included feature and are not shown. The Shapley transformation deductively infers the local effect of each feature with respect to local outcome. These evaluations are internally independent, case-to-case, resulting in a variability profile markedly different from the original data. For example, low-resolution features that are homogenous region to region are still likely to be assigned unique values by the Shapley transformation. The SHAP-derived effect matrix was found to outperform the original, nominal value data in multivariate prediction, even when decomposed according to feature class. The magnitude of performance increase was most pronounced for the feature classes comprised of lower-resolution data. However, even in the case of climate, the benefit of SHAP is undeniable particularly with smaller feature set sizes.

[*Legend F2b]*

SHAP model decomposition reveals domain-specific patterns of outbreak risk. The ability to arbitrarily decompose a given prediction presents powerful new opportunities. Risk maps were generated according to feature class and rescaled for visual clarity; values are summative with respect to the overall likelihood estimates of West Nile virus outbreak in 2018. The differential effect of the each feature class was found to vary substantially as a function of geography and scale. The effect of climate was confined mostly to the southern latitudes with some notable spread into France and Germany. Risk associated with the environmental feature class had a notable western orientation with homogenous degrees of risk extending from Greece in the south to Sweden in the north. Outbreak risk associated with economic disparities was markedly widespread across, with the Scandinavian countries and UK being among the most notable exceptions. Presence and diversity of potential hosts contributed most notably to outbreak risk in Spain. As for known and potential mosquito vectors, our results showed a far more sporadic pattern of risk. In addition to being substantial in the WNV endemic regions of Italy and Romania and neighboring, vector-associated risk was identified almost uniformly throughout Spain and even into parts of Sweden, Denmark, and nearby Germany.

[Table 1]

Parameter estimates and R-squared results from linear regression. Significant parameters and respectable goodness of fit confirms empirical robustness of SHAP effect estimates.

References

1. Patel, H., Sander, B. & Nelder, M. P. Long-term sequelae of West Nile virus-related illness: a systematic review. *Lancet Infect. Dis.* **15**, 951–959 (2015).
2. Hernández-Triana, L. M. *et al.* Emergence of West Nile Virus Lineage 2 in Europe: A Review on the Introduction and Spread of a Mosquito-Borne Disease. *Front. Public Health* **2**, (2014).
3. Camp, J. V. & Nowotny, N. The knowns and unknowns of West Nile virus in Europe: what did we learn from the 2018 outbreak? *Expert Rev. Anti Infect. Ther.* **18**, 145–154 (2020).
4. Keeling, M. J. & Rohani, P. *Modeling Infectious Diseases in Humans and Animals*. (Princeton University Press, 2011).
5. Dodd, R. Y., Foster, G. A. & Stramer, S. L. Keeping Blood Transfusion Safe From West Nile Virus: American Red Cross Experience, 2003 to 2012. *Transfus. Med. Rev.* **29**, 153–161 (2015).
6. Anesi, J. A. & Silveira, F. P. Arenaviruses and West Nile Virus in solid organ transplant recipients: Guidelines from the American Society of Transplantation Infectious Diseases Community of Practice. *Clin. Transplant.* **33**, e13576 (2019).
7. Centers for Disease Control and Prevention (CDC). West Nile virus infection among turkey breeder farm workers--Wisconsin, 2002. *MMWR Morb. Mortal. Wkly. Rep.* **52**, 1017–1019 (2003).
8. Centers for Disease Control and Prevention (CDC). Laboratory-acquired West Nile virus infections--United States, 2002. *MMWR Morb. Mortal. Wkly. Rep.* **51**, 1133–1135 (2002).
9. Hayes, E. B. & O'Leary, D. R. West Nile Virus Infection: A Pediatric Perspective. *Pediatrics* **113**, 1375–1381 (2004).
10. Chianese, A. *et al.* West Nile virus: an overview of current information. *Transl. Med. Rep.* **3**, (2019).
11. Pisani, G., Cristiano, K., Pupella, S. & Liumbruno, G. M. West Nile Virus in Europe and Safety of Blood Transfusion. *Transfus. Med. Hemotherapy* **43**, 158–167 (2016).
12. Mrzljak, A. *et al.* West Nile Virus: An Emerging Threat in Transplant Population. *Vector-Borne Zoonotic Dis.* (2020) doi:10.1089/vbz.2019.2608.

13. Pachler, K. *et al.* Putative new West Nile virus lineage in Uranotaenia unguiculata mosquitoes, Austria, 2013. *Emerg. Infect. Dis.* **20**, 2119 (2014).
14. Vázquez, A. *et al.* Putative new lineage of West Nile virus, Spain. *Emerg. Infect. Dis.* **16**, 549 (2010).
15. Bondre, V. P., Jadi, R. S., Mishra, A. C., Yergolkar, P. N. & Arankalle, V. A. West Nile virus isolates from India: evidence for a distinct genetic lineage. *J. Gen. Virol.* **88**, 875–884 (2007).
16. Shahhosseini, N. & Chinikar, S. Genetic evidence for circulation of Kunjin-related West Nile virus strain in Iran. *J. Vector Borne Dis.* **53**, 384 (2016).
17. McMullen, A. R. *et al.* Molecular evolution of lineage 2 West Nile virus. *J. Gen. Virol.* **94**, 318–325 (2013).
18. Ruiz, M. O., Tedesco, C., McTighe, T. J., Austin, C. & Kitron, U. Environmental and social determinants of human risk during a West Nile virus outbreak in the greater Chicago area, 2002. *Int. J. Health Geogr.* **3**, 8 (2004).
19. Davis, J. K. *et al.* Improving the prediction of arbovirus outbreaks: A comparison of climate-driven models for West Nile virus in an endemic region of the United States. *Acta Trop.* **185**, 242–250 (2018).
20. Semenza, J. C. Prototype Early Warning Systems for Vector-Borne Diseases in Europe. *Int. J. Environ. Res. Public. Health* **12**, 6333–6351 (2015).
21. Barker, C. M. Models and Surveillance Systems to Detect and Predict West Nile Virus Outbreaks. *J. Med. Entomol.* **56**, 1508–1515 (2019).
22. Elith, J., Leathwick, J. R. & Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **77**, 802–813 (2008).
23. Lundberg, S. M. *et al.* From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 56–67 (2020).
24. Chen, T. & Guestrin, C. Xgboost: A scalable tree boosting system. in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* 785–794 (2016).

25. Lundberg, S. M., Erion, G. G. & Lee, S.-I. Consistent Individualized Feature Attribution for Tree Ensembles. *ArXiv180203888 Cs Stat* (2019).
26. Munkhdalai, L., Munkhdalai, T., Namsrai, O.-E., Lee, J. & Ryu, K. An Empirical Comparison of Machine-Learning Methods on Bank Client Credit Assessments. *Sustainability* **11**, 699 (2019).
27. He, X. *et al.* Practical Lessons from Predicting Clicks on Ads at Facebook. in *Proceedings of 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining - ADKDD'14* 1–9 (ACM Press, 2014). doi:10.1145/2648584.2648589.
28. Nielsen, D. Tree boosting with xgboost-why does xgboost win" every" machine learning competition? (NTNU, 2016).
29. Shapley, L. S. Notes on the n-Person Game—II: The Value of an n-Person Game. (1951).
30. Simplify the acquisition of remote sensing data. https://jesjehle.github.io/earthEngineGrabR/.
31. IRI Climate and Society Map Room. index.html.
32. Kriticos, D. J., Jarošik, V. & Ota, N. Extending the suite of BIOCLIM variables: a proposed registry system and case study using principal components analysis. *Methods Ecol. Evol.* **5**, 956–960 (2014).
33. Hijmans, R. J., Phillips, S. & Elith, J. L. and J. *dismo: Species Distribution Modeling*. (2017).
34. Import Various Northern and Southern Hemisphere Climate Indices. https://boshek.github.io/rsoi/.
35. Keyel, A. C. *et al.* Seasonal temperatures and hydrological conditions improve the prediction of West Nile virus infection rates in Culex mosquitoes and human case counts in New York and Connecticut. *PLOS ONE* **14**, e0217854 (2019).
36. European Environment Agency's home page — European Environment Agency. https://www.eea.europa.eu/.
37. Natura 2000 - Environment - European Commission. https://ec.europa.eu/environment/nature/natura2000/index_en.htm.
38. Lahti [aut, L. *et al. eurostat: Tools for Eurostat Open Data*. (2020).

39. Laperriere, V., Brugger, K. & Rubel, F. Simulation of the seasonal cycles of bird, equine and human West Nile virus cases. *Prev. Vet. Med.* **98**, 99–110 (2011).
40. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
41. Wang, Y. & Ni, X. S. A XGBoost risk model via feature selection and Bayesian hyper-parameter optimization. *ArXiv190108433 Cs Stat* (2019).
42. Xia, Y., Liu, C., Li, Y. & Liu, N. A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Syst. Appl.* **78**, 225–241 (2017).
43. Mayr, A., Binder, H., Gefeller, O. & Schmid, M. The evolution of boosting algorithms. From machine learning to statistical modelling. *Methods Inf. Med.* **53**, 419 (2014).
44. Blagus, R. & Lusa, L. Gradient boosting for high-dimensional prediction of rare events. *Comput. Stat. Data Anal.* **113**, 19–37 (2017).
45. Cortes, C. & Mohri, M. AUC optimization vs. error rate minimization. in *Advances in neural information processing systems* 313–320 (2004).
46. Yang, K., Yu, R., Wang, X., Quddus, M. & Xue, L. How to determine an optimal threshold to classify real-time crash-prone traffic conditions? *Accid. Anal. Prev.* **117**, 250–261 (2018).
47. XGBoost. https://xgboost.ai/.
48. Lefcheck, J. S. piecewiseSEM: Piecewise structural equation modelling in r for ecology, evolution, and systematics. *Methods Ecol. Evol.* **7**, 573–579 (2016).
49. Wickham, H. *et al. ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. (2020).
50. Tennekes, M. *et al. tmap: Thematic Maps*. (2020).
51. Dowle, M. *et al. data.table: Extension of 'data.frame'*. (2019)