

BEXYL
BEYOND XYLELLA

Maps of relative risk to *Xylella fastidiosa* for selected areas in the EU based on the identified major biotic and abiotic factors that drive *X. fastidiosa* epidemics

Deliverable 1.1

30 April 2026

Author(s): Juan A. Navas-Cortés, Blanca B. Landa, Luis F. Arias-Giraldo Rocío Calderón (P1-CSIC), Leonardo de la Fuente (P9-AU)

Funded by the European Union

DELIVERABLE INFORMATION

Start of the project:	01 Sep 2022
Duration:	48 months
Coordinator:	Blanca B. Landa
Deliverable title:	Maps of relative risk to *Xylella fastidiosa* for selected areas in the EU based on the identified major biotic and abiotic factors that drive *X. fastidiosa* epidemics
Deliverable n°:	1.1
Nature of the deliverable:	Other
Dissemination level:	Public
WP contributing:	WP1
Lead beneficiary:	CSIC
Due date of deliverable:	30 April 2026
Actual submission date:	30 April 2026

Deliverable status

Ver.	Status	Date	Author(s)
1.0	Final	30 April 2026	Juan A. Navas-Cortés, Blanca B. Landa, Luis F. Arias-Giraldo, Rocío Calderón (P1-CSIC), Leonardo de la Fuente (P9-AU)

Disclaimer

The content of this deliverable does not necessarily reflect the official opinions of the European Commission or other institutions of the European Union.

Funded by the European Union

Table of contents

1. Risk maps for *X. fastidiosa* based on climatic variables....5
1.1. Occurrence data....5
1.2. Climatic data....5
1.3. Modelling region definition....5
1.4. Algorithm-specific implementation....5
1.5. Variable selection....5
1.6. Model training and evaluation....6
1.7. Algorithm-specific model implementation....6
1.8. Model performance metrics and evaluation....6
1.9. Algorithm-specific variable importance extraction....6
1.10. Climatic risk maps....6
2. Risk maps for *X. fastidiosa* based on soil edaphic properties....7
2.1. Occurrence data....7
2.2. Soil edaphic properties data....7
2.3. Algorithm implementation....7
2.4. Model training and evaluation....7
2.5. Soil risk maps....8
3. References....9

Funded by the European Union

BACKGROUND

Xylella fastidiosa represents one of the most significant emerging plant pathogen threats to European agriculture and natural ecosystems. This gram-negative, xylem-limited bacterium has caused devastating outbreaks across multiple continents, with recent introductions into Europe highlighting the urgent need for comprehensive risk assessment and management strategies. The impact of the pathogen impact extends beyond agricultural crops to include ornamental plants and forest species, making it a critical concern for both economic and ecological sustainability in the European Union.

The emergence of *X. fastidiosa* in Europe, particularly the outbreaks in southern Italy affecting olive trees and the subsequent detections across the Mediterranean basin, has demonstrated its potential for rapid spread and severe economic impact. The complex epidemiology of *X. fastidiosa* involves intricate interactions between the pathogen, its host plants, insect vectors, and environmental factors including climate, soil conditions, and landscape composition. Understanding these multifactorial relationships is essential for developing effective surveillance, containment, and management strategies.

Previous European research initiatives, including POnTE and XF-ACTORS, have established foundational knowledge about *X. fastidiosa* biology, host range, and climatic suitability in European contexts. However, significant knowledge gaps remain regarding the specific biotic and abiotic factors that drive epidemic development and the spatial risk across different European landscapes and agricultural systems.

Risk mapping represents a critical tool for proactive management of *X. fastidiosa* threats, enabling targeted surveillance efforts, informed regulatory decisions, and strategic allocation of resources for prevention and control measures. The development of spatially explicit risk models requires integration of comprehensive epidemiological data including environmental variables.

This deliverable (D1.1) addresses these needs by developing maps of relative risk to *X. fastidiosa* for the EU, building upon the comprehensive understanding of major biotic and abiotic drivers identified through Tasks 1.1 and 1.2 of the BeXyl project. The risk assessment framework incorporates field survey data, laboratory studies on host-pathogen interactions under various environmental conditions, and advanced spatial modeling approaches. These risk maps provide tools for EU plant health authorities, enabling evidence-based decision-making for *X. fastidiosa* surveillance, quarantine measures, and long-term management strategies.

During the project progress since D1.1 submission until the end of the project the maps will be updated if new occurrences data become available from the different sources indicated in section 1.1.

Funded by the European Union

1. RISK MAPS FOR *X. fastidiosa* BASED ON CLIMATIC VARIABLES

1.1. Occurrence data

Occurrence data for *X. fastidiosa* were compiled from multiple complementary sources to ensure broad geographic coverage and data reliability. These sources include: (1) the Update of the *Xylella* spp. Host database (EFSA et al., 2026); (2) national and regional outbreak datasets provided by National Plant Protection Organizations from Italy, France, Portugal, and Spain; (3) records from the Hula Valley, Israel, reported in the EPPO Reporting Service (2019/121); and (4) peer-reviewed scientific literature reporting georeferenced occurrences.

1.2. Climatic data

Climatic data was obtained from CHELSA climatologies v2.1 (Krager et al. 2017) for preselected climatic variables

1.3 Modelling region definition

Geographic constraints: Pseudo-absences are constrained by:

(1) land-water boundaries using high-resolution masks,

(2) elevation limits excluding areas >2,500m elevation that exceed physiological tolerances

(3) latitudinal bounds (60°N to 45°S) reflecting thermal tolerance limits

(4) minimum distance buffers (0.1°) from presence points, and

(5) minimum distance buffers (0.1°) from pseudoabsence points.

1.4 Algorithm-specific implementation

Optimal pseudo-absence strategies vary significantly among modeling algorithms. Three distinct algorithm groups with different pseudo-absence requirements were implemented: (i) Regression methods (GLM, GAM, GLMNET): Show little sensitivity to pseudo-absence number but benefit from large samples (~10,000) with equal weighting one replication each; (ii) Classification methods (BRT, Random Forest): Achieve optimal performance with 1:1 presence-to-pseudoabsence ratios with 10 replications each; (iii) Intermediate methods (MARS): Perform well with moderate pseudo-absence numbers (~100) with 10 replications (Barbet-Massin et al., 2012).

1.5 Variable selection

Variable selection proceeded through four complementary approaches: pairwise correlation analysis, Variance Inflation Factor (VIF) analysis, hierarchical clustering, and Principal Components Analysis (PCA), with optional multi-resolution assessment.

Funded by the European Union

1.6. Model training and evaluation

For model training we used pre-extracted environmental datasets generated by model-specific pseudo-absence procedures. Spatial autocorrelation in species occurrence data was addressed through spatial block cross-validation. Block size calculation was adapted to dataset spatial extent ensuring adequate separation between training and testing folds (Valavi et al., 2019).

1.7. Algorithm-specific model implementation

Six modeling algorithms were implemented with algorithm-specific parameter optimization:

(1) GLM (Generalized Linear Models) with binomial family and logit link with all environmental variables as linear predictors; (2) GAM (Generalized Additive Models) with smooth terms for each variable, and REML optimization; (3) Random Forest implemented with classification approach, dynamic parameters including sample-size dependent parameters, and class balancing calculated as ratio of absence to presence counts; (4) GLMNET (Regularized Regression) with cross-validated lambda selection, LASSO regularization, 10-fold CV; (5) MARS (Multivariate Adaptive Regression Splines) with GLM binomial family for binary response

1.8. Model performance metrics and evaluation

Six performance metrics were implemented with algorithm-specific prediction methods: AUC (Area Under the Curve), TSS (True Skill Statistic), Miller Calibration Slope, Sensitivity and Specificity at TSS-optimal threshold, Kappa Statistic (Allouche et al. 2006; Liu et al., 2013).

Two evaluation procedures assessed model performance at different levels: (1) Training evaluation assessed models on their training datasets to evaluate fit quality and potential overfitting, with predictions generated using algorithm-specific methods with appropriate parameters. (2) Spatial Cross-Validation evaluation implemented independent model fitting and evaluation for each spatial fold.

1.9. Algorithm-specific variable importance extraction

Variable importance extraction utilized algorithm-specific methods reflecting different statistical foundations: (1) Classical algorithm-specific importance utilized native extraction methods for each algorithm; (2) Shapley values (SHAP) quantified model-agnostic variable contributions.

1.10. Climatic risk maps

The climatic risk maps are available for download at the official SACO-CSIC cloud of BeXyl project: https://saco.csic.es/s/4PzgWCw6cW7ERrH

During the period from the submission of D1.1 until the end of the project, the maps will be updated whenever new occurrence data become available from the different sources indicated in Section 1.1.

Funded by the European Union

2. RISK MAPS FOR X. *fastidiosa* BASED ON SOIL EDAPHIC PROPERTIES

2.1. Occurrence data

Occurrence data for *X. fastidiosa* were compiled from multiple complementary sources to ensure broad geographic coverage and data reliability. These sources include: (1) the Update of the *Xylella* spp. Host database (EFSA et al., 2026); (2) national and regional outbreak datasets provided by National Plant Protection Organizations from Italy, France, Portugal, and Spain; (3) records from the Hula Valley, Israel, reported in the EPPO Reporting Service (2019/121); and (4) peer-reviewed scientific literature reporting georeferenced occurrences.

2.2. Soil edaphic properties data

Soil edaphic data were obtained from the SoilGrids v2.0 global database (Poglio et al., 2020) and complemented with categorical soil information from FAO soil datasets. Both continuous and categorical soil variables were considered.

Continuous soil variables included bulk density of the fine earth fraction (BDOD, kg/dm^3), cation exchange capacity (CEC, cmol(c)/kg), volumetric fraction of coarse fragments (> 2 mm; CFVO, vol%), soil texture components; clay, sand, and silt content expressed as proportions of the fine earth fraction (g/100 g, %); soil organic carbon content (SOC, g/kg), total nitrogen content (N, g/kg), and soil pH (pH units x 10). Organic carbon density (OCD, kg/m^3) and organic carbon stocks (OCS, kg/m^2) were also considered where relevant. All variables were converted from their native mapped units to conventional units using standardized conversion factors. Soil layers were extracted at standard reference depths and resampled to a common spatial resolution.

2.3. Algorithm implementation

The potential distribution ranges for *X. fastidiosa* were modeled using a multi-ensemble approach that included six algorithms implemented in the Biomod2 R package (Thuiller et al., 2009): Boosted Regression Trees (BRT), Classification Tree Analysis (CTA), Artificial Neural Networks (ANN), Multivariate Adaptive Regression Splines (MARS), Random Forest (RF), and Maximum Entropy (MaxEnt). Algorithm-specific parameter optimization was applied following BIOMOD2 default and recommended practices.

To reduce spatial autocorrelation and sampling bias, presence points were spatially thinned (Aiello-Lammens et al., 2015) to a minimum distance of 0.083° grid resolution (~9.2 km at the equator). Pseudo-absences were generated to calibrate the model using used a combined random and buffer-exclusion approach as described before (Barbet-Massin et al. 2012).

2.4. Model training and evaluation

Models were calibrated using the presence and pseudo-absence datasets described above. Model evaluation was conducted using repeated k-fold cross-validation with 5 folds (k = 5) and 4 repetitions. Predictive performance was primarily assessed using the True Skill Statistic (TSS), and only models meeting predefined performance criteria were retained for ensemble

Funded by the European Union

construction. Final soil-based risk maps were produced as weighted ensemble predictions, with individual model contributions weighted according to their TSS values (Allouche et al., 2006).

Model performance was further evaluated using the same set of metrics applied in the climatic modelling framework, including Area Under the Curve (AUC), TSS, Kappa statistic, sensitivity and specificity at the TSS-optimal threshold, and calibration metrics. Performance was assessed both on training datasets and under spatial cross-validation.

2.5. Soil risk maps

The soil risk maps are available for download at the official SACO-CSIC cloud of BeXyl project: https://saco.csic.es/s/ZRpCij2s32MGDwb

During the period from the submission of D1.1 until the end of the project, the maps will be updated whenever new occurrence data become available from the different sources indicated in Section 1.1.

Funded by the European Union

3. REFERENCES

Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. Ecography, 38(5), 541-545.

Allouche O, Tsoar A, Kadmon R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology 43:1223-32

Barbet-Massin, M., Jiguet, F., Albert, C.H., Thuiller, W. 2012. Selecting pseudo-absences for species distribution models: how, where and how many? Methods in Ecology and Evolution, 3(2), 327-338.

EFSA, Cavalieri, V., Fasanelli, E., Furnari, G., Gibin, D., Gutierrez Linares, A., La Notte, P., Pasinato, L., Stancanelli, G. 2026. Update of the *Xylella* spp. host plant database – Systematic literature search up to 30 June 2025. EFSA Journal 24:e9840

Karger, D.N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R.W., Zimmermann, N.E., Linder, P., Kessler, M. 2017. Climatologies at high resolution for the Earth's land surface areas. Scientific Data, 4, 170122.

Liu, C., White, M., Newell, G. 2013. Selecting thresholds for the prediction of species occurrence with presence-only data. Journal of Biogeography, 40(4), 778-789.

Poggio, L., De Sousa, L. M., Batjes, N. H., Heuvelink, G., Kempen, B., Ribeiro, E., & Rossiter, D. (2021). SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. Soil, 7(1), 217-240.

Thuiller, W., Lafourcade, B., Engler, R., & Araújo, M. B. (2009). BIOMOD–a platform for ensemble forecasting of species distributions. Ecography, 32(3), 369-373.

Valavi, R., Elith, J., Lahoz-Monfort, J.J., Guillera-Arroita, G. 2019. blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. Methods in Ecology and Evolution, 10(2), 225-232.

Funded by the European Union