Application of text mining and environmental data for the assessment of West Nile risk in Europe

Elena Arsevska, UMR ASTRE, CIRAD

18 October 2025

Few words about me and the funding of this work

- Researcher in vet epidemiology
- One Health topics - transboundary & zoonotic infectious diseases
- Epidemiological analysis of data from complementary & non-traditional sources for improved epidemic intelligence (EI)
- Support to health (PH/AH) agencies – knowledge transfer and capacity building

- MOnitoring Outbreak events for Disease surveillance in a data science context (MOOD)
- H2020 (2020-2024) research & innovation (23 partners)
- Co-develop with EI practitioners tools and services for improved detection, monitoring and assessment of disease outbreaks in Europe by assembling multi-source data
- Deliver a freely accessible visual dashboard (MOOD platform) to EI practitioners and contributors
- Several case studies (vector-borne, respiratory pathogens) of importance to Europe

Epidemic intelligence (EI) in theory

All activities related to the early identification of potential health hazards, their investigation and verification, assessment, and communication *(Paquet et al.,2006)*

Figure source: European Commission

Epidemic intelligence practitioners and contributors

- Epidemiologists at public health and animal health agencies (PH/AH)
  - Monitor disease threats at national/international level
  - Prepare surveillance and control plans
- Risk assessors
  - Assess the risk of the introduction or spread of pathogens
- Decision makers
  - Need to set appropriate mitigation measures
- Researchers
  - Range of domains (informatics, epidemiology, data science and modelling)
  - Methodological and technical solutions

Epidemic intelligence ideal situation for EI practitioners

Need of timely, available, accessible & quality data for efficient risk assessment, preparedness and response

TESSy

THE NEW OIE-WAHIS World Animal Health Information System

PADI - WEB

EIOS EPIDEMIC INTELLIGENCE FROM OPEN SOURCES

PubMed®

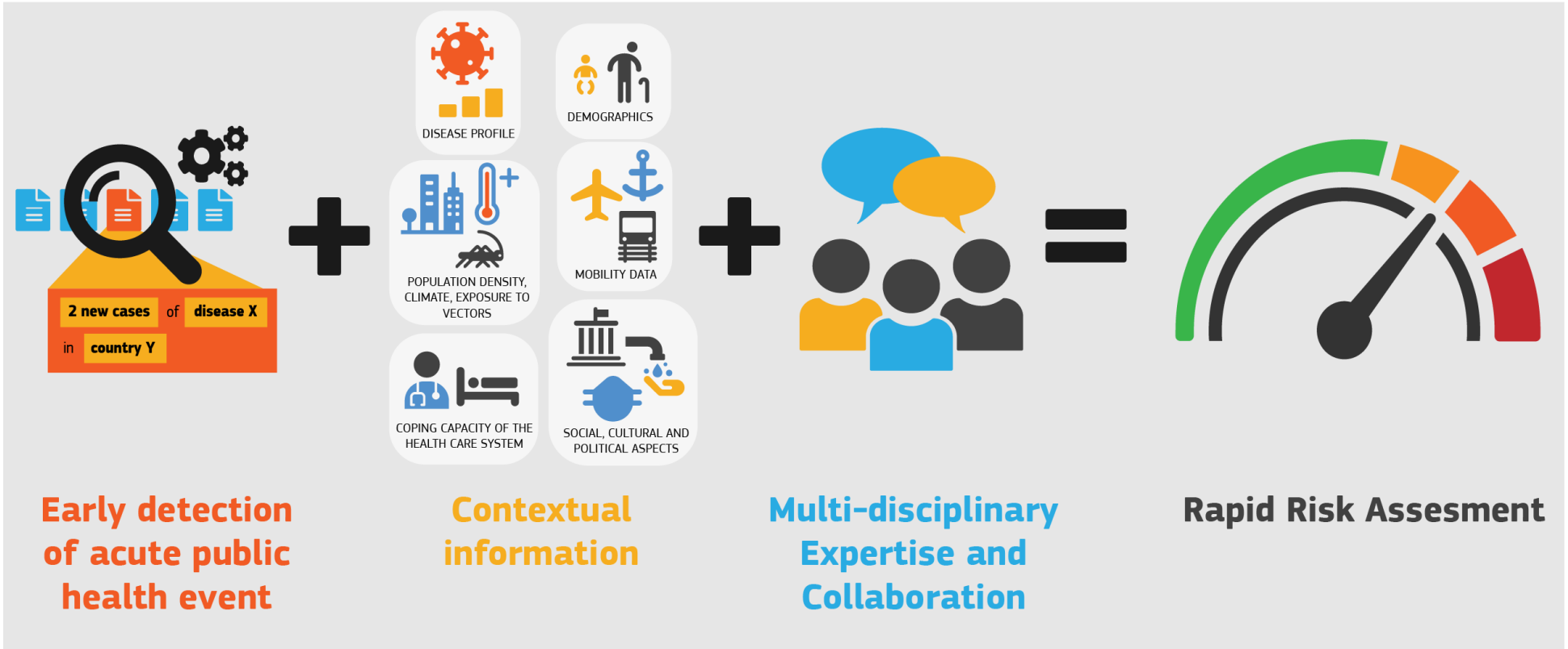


Figure source: European Commission

Epidemic intelligence challenges – West Nile virus in Europe

Not timely, difficult to find, not accessible & low resolution data for risk assessment

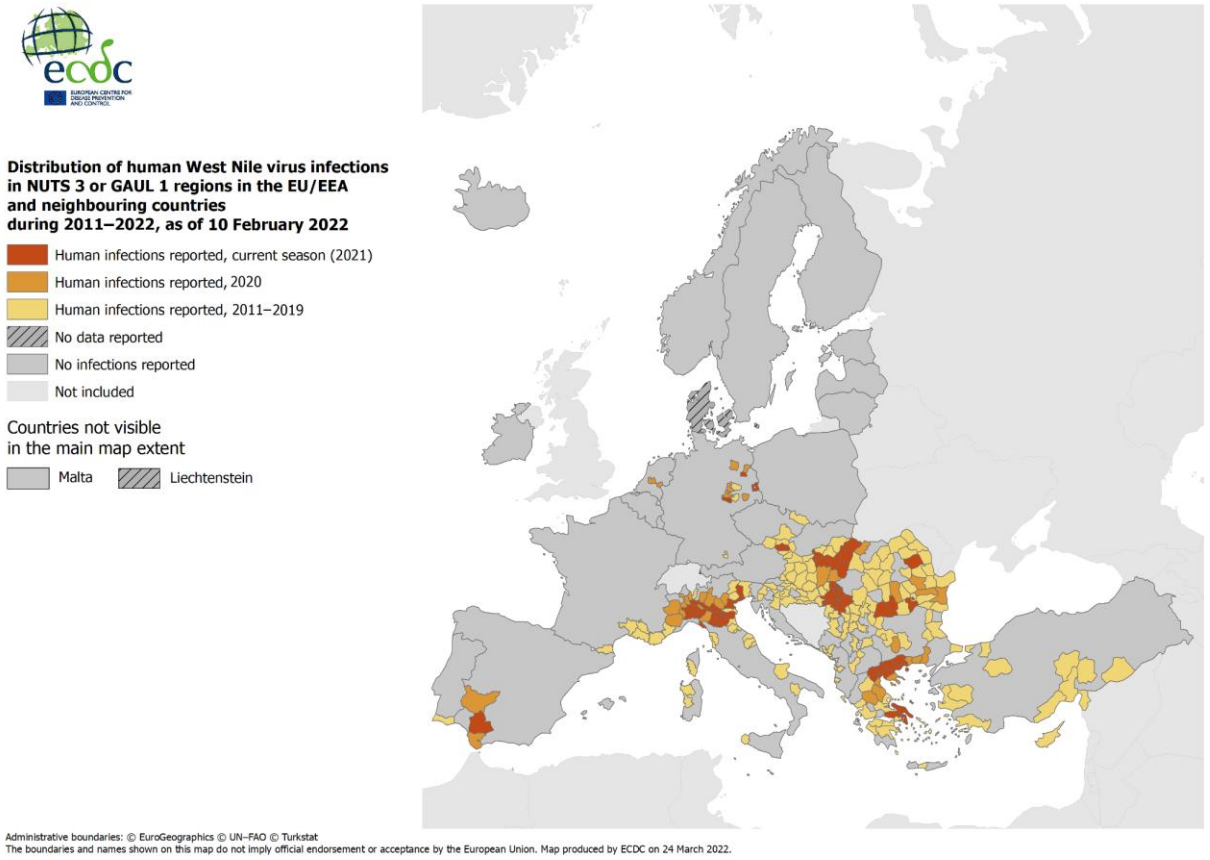


Figure source: ECDC

- Vector-borne disease (birds, horses, humans)
- Notifiable disease (since 2008 to ECDC) and WOAH
- Progressive endemicity (annual cases observed since 2010)
- Seasonal transmission → linked to the vector activity period (April to November)
- Emergence influenced by various climatic factors
- Surveillance based on key indicators is important (weekly maps + annual data analysis by the ECDC)
- Reporting at different NUTS levels (Nomenclature of Territorial Units for Statistics) depending on the country → NUTS 3 highest (best resolution)

Objectives of the work

- Asses how event-based surveillance data from informal sources **EBS** (e.g., PADI-web) **can complement** indicator-based data from traditional surveillance **IBS** (e.g., ECDC TESSy) to enhance risk mapping of VBD in Europe, using WNV as a case study

- We used both EBS and IBS data to train **two ecological niche models** using a boosted regression trees (**BRT**) approach (presence/absence; binomial) to estimate the ecological suitability of WNV given local environmental conditions per NUTS3 level
  - **Model 1: IBS data only and**
  - **Model 2: IBS and EBS**
  - **100 replicates for each model and block cross-validation for the pseudo-absence NUTS3 levels**

- Provide **FAIR** data, code, model outputs and materials for use by **national health authorities/ ECDC and researchers though the MOOD platform**

Methods – epi data

- IBS data from ECDC TESSy on autochthonous WNV in humans from **2006 to 2021** from 182 NUTS3 (**data upon request for research only**)
- EBS data on WNV in humans from **2006 to 2021** from PADI-web tool (**freely available data**)
  - Set queries to search Google news in 16 languages based on countries with previous WNV
  - Retrieve news (n=706) on WNV cases
  - Verify and retain news reporting at least one autochthonous WNV human case (n=250)
  - Extract locations of a human case (n=2356)
  - Verify and retain relevant locations (n=766 + 192 added because the system did not extract)
  - Aggregate at NUTS3 level for modelling
  - Share dataset (raw and clean) in Dataverse

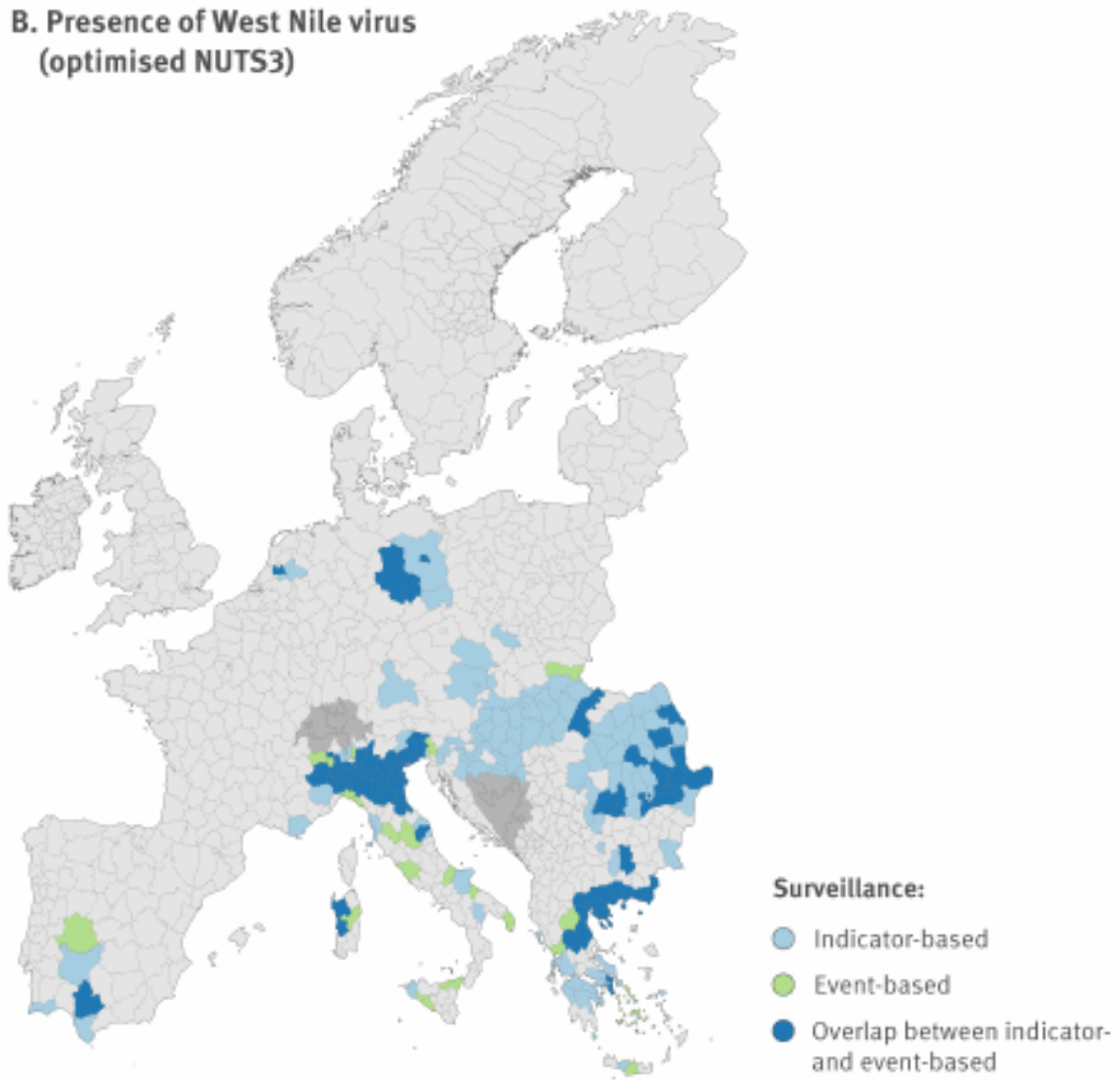


Figure source: Serres et al, 2024

Methods – spatio-temporal covariate data

14 environmental factors documented to influence the risk of circulation of WNV in Europe

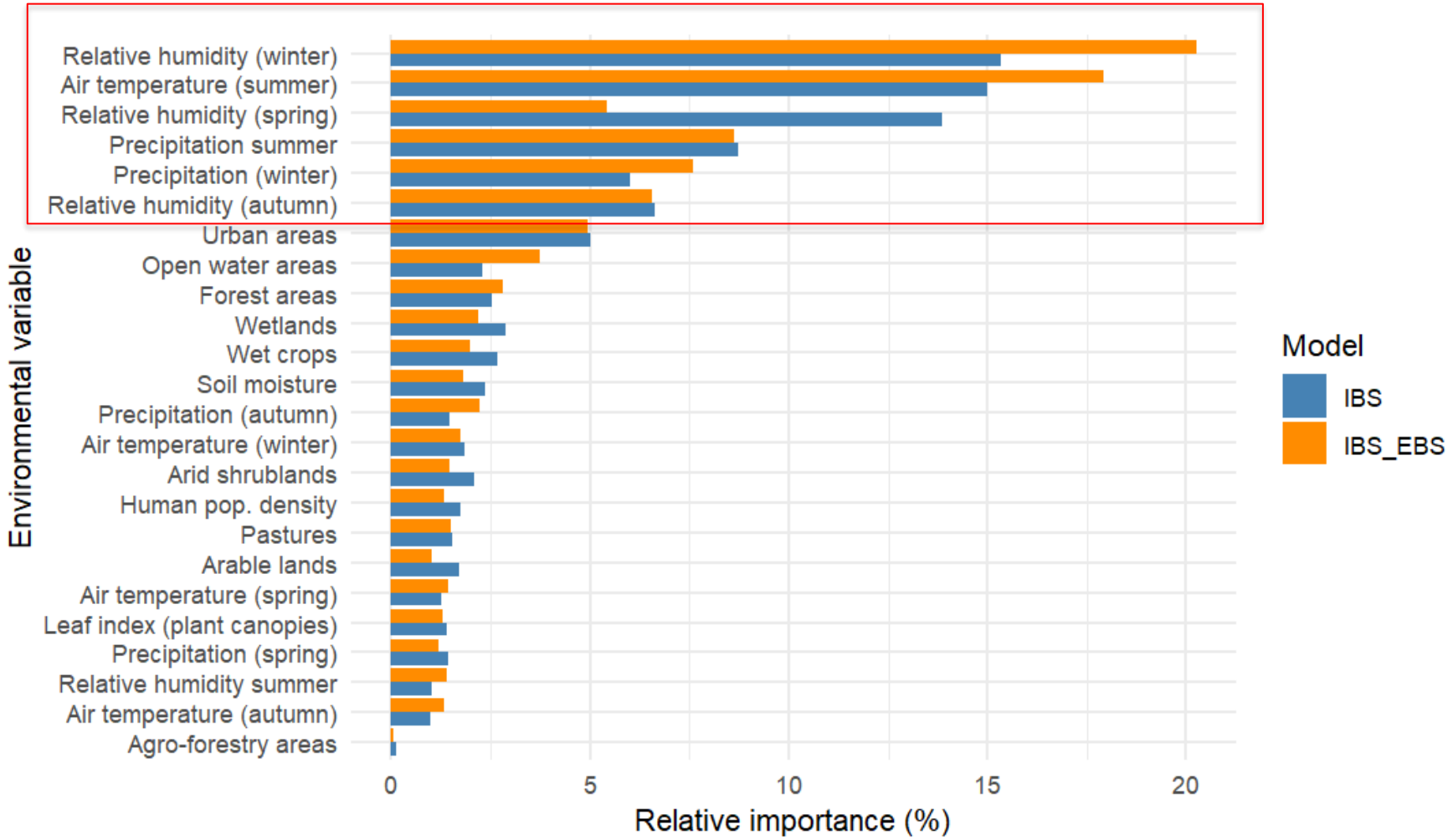
Environmental factors	Variables used in model	Unit of measurment	Spatial resolution	Temporal resolution	Time period	Source	DOI
Human population[1, 5, 23]	Human population	log10	0.5°	yearly	2000-2021	ISIMIP	https://doi.
Mean air temperature[1-4, 7-41, 43, 44, 46]	2-meter air temperature (by season)	Celcius	0.25°	monthly	2000-2021	Era5	https://doi.
Precipitation[1-2, 4, 7, 11-16, 22, 24-28, 32-41, 43, 44]	Total precipitation (by season)	cm					
Relative humidity[1, 5, 9, 33, 39, 40]	Relative humidity (by season)	%					https://doi.
Soil moisture[7, 20, 35, 37, 42]	Volumetric soil water	m^3					https://doi.
Vegetation[3, 8, 14, 16, 20, 24, 26-27, 35, 37, 42]	Leaf index (plant canopies)	%	100m	yearly	2000-2006 and 2012-2018	CLC	https://doi.
Urban areas[1, 4, 7, 20, 24, 27, 36-37, 41-42, 44-45]	Urban areas						
Agriculture[1, 4-7, 10-11, 20, 24, 27, 30, 36-37, 39, 41-42, 44-45]	Wet crops Pastures Agro-forestry areas						
Forest areas[1, 4-7, 10-11, 20, 24, 27, 30, 36-37, 39, 41-42, 44-45]	Forest areas						
Rural areas[1, 4, 7, 20, 24, 27, 36-37, 41-42, 44-45]	Arable lands						
	Aird shrublands						
Wetlands[1, 4-7, 10-11, 20, 24, 27, 30, 36-37, 39, 41-42, 44-45]	Wetlands						
Aquatic habitat[2, 4, 35, 42]	Open water areas						

Environmental Research Group Oxford

Table source: Serres et al, 2024

Results – influence covariates on WNV occurrence

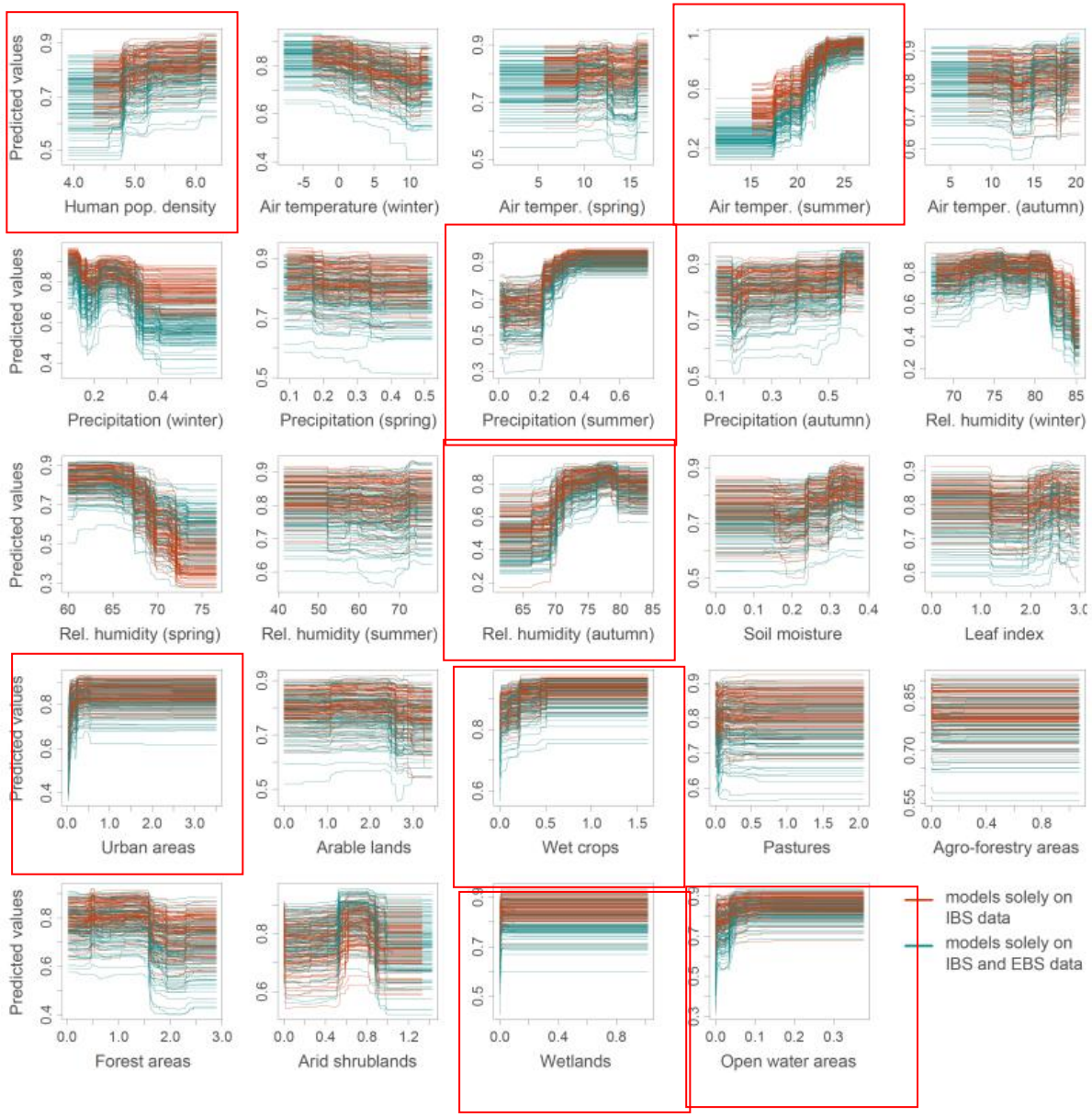
Climatic factors (**temperature, precipitation & humidity**), exhibit the highest influence (RI >10%) on the probability of WNV human infection in Europe in **both models**, i.e. either trained on **IBS and IBS+EBS data**



Results – association covariates vs WNV occurrence

Response curves show of **both models** (model IBS and IBS+EBS) that **temperatures** ranging from 15 to 28° C, **precipitation** from 0.2 to 0.5 cm and **relative humidity** levels from 70 to 85% had a **positive influence** on the probability of the occurrence of human cases of WNV

Both models (model IBS and IBS+EBS) also showed clear **positive associations** between ecological suitability for local WNV circulation and **human population density, urban areas, open water areas, wet crops and wetlands**, as well as the level of soil moisture



Results – probability of WNV occurrence

Both models (IBS and IBS+EBS) identified highly suitable regions for local WNV circulation leading to human cases, particularly in **southern Iberian Peninsula, southern France, Italy, the Balkan Peninsula and countries in the Danube River Basin**

Model2: IBS+EBS models identified additional areas potentially at risk of local WNV circulation, specifically in **central and southern Italy, southern Spain, southern France and Greece**

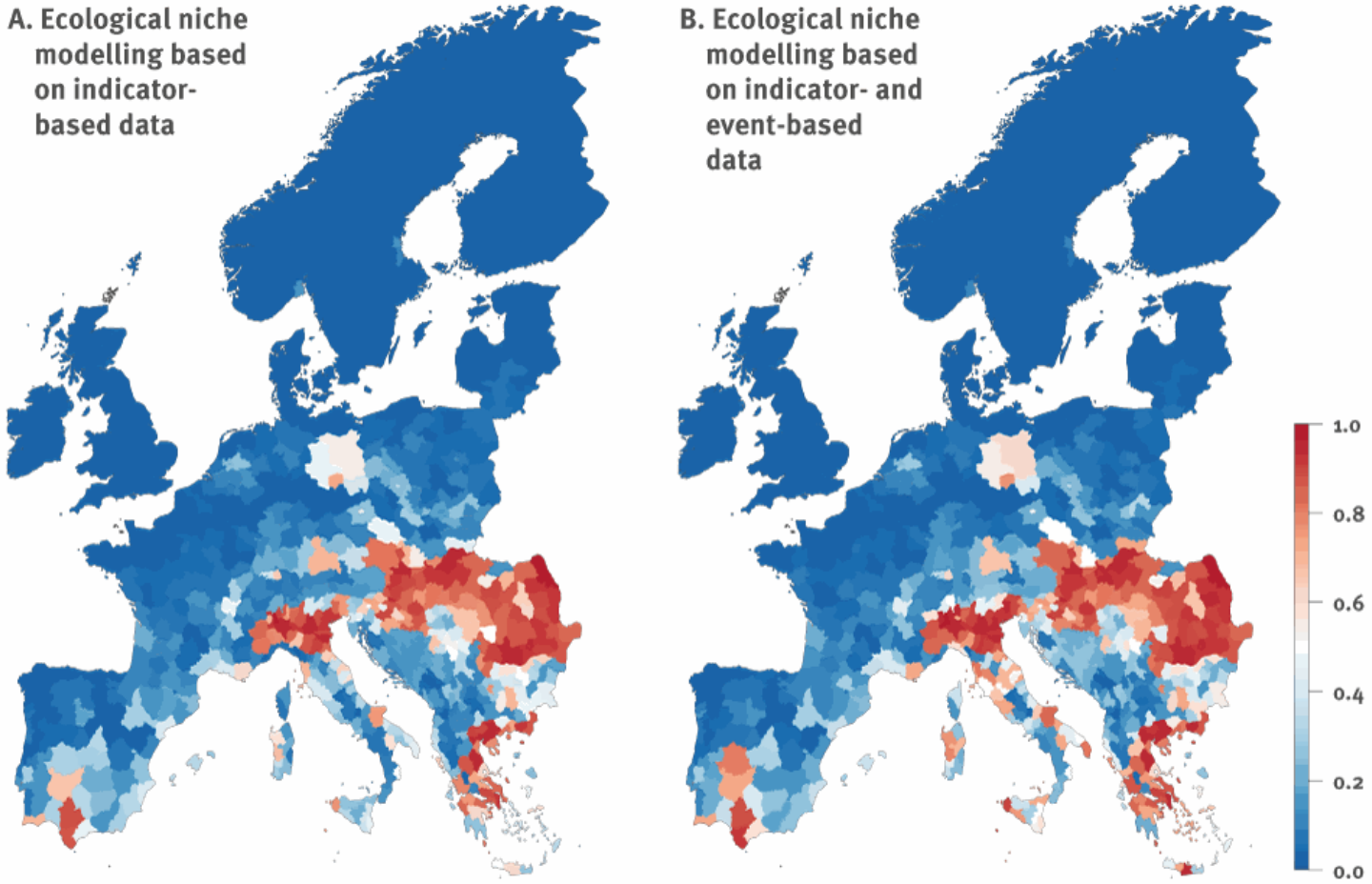


Figure source: Serres et al, 2024

Training

MOOD

MOnitoring Outbreaks for Disease surveillance in a data science context

SUMMER SCHOOL 2023

Organized by: OpenGeoHUB

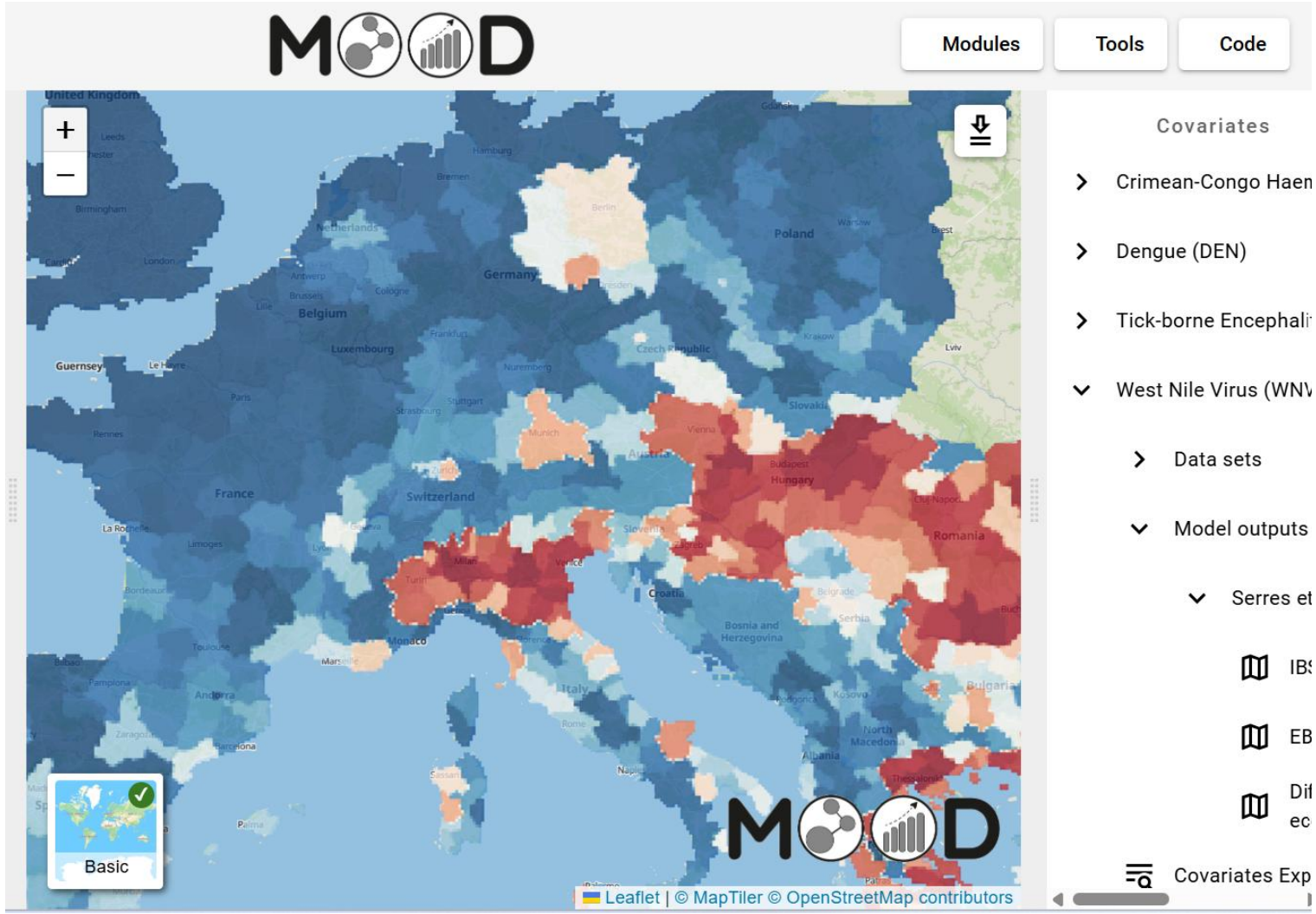
May 21st - 27th 2023

The Mediterranean Institute for Life Sciences, MedILS, Split, Croatia

OpenGeoHUB
Connect • Create • Share • Repeat

Watch all demos (data, code & tutorials) HERE

Model implementation in the MOOD platform



MOOD platform freely accessible: https://app.mood-h2020.eu/core?tab=diseases

Summary

What have we learnt from this study?

- By using environmental data and two types of surveillance data (IBS & EBS), we identified new areas in southern Europe potentially suitable for local circulation of WNV leading to human cases
- This work would not be possible without a pluri-disciplinary approach and involvement of computer science, ecology, epidemiology, EI practitioners and future users of the results

Implications for EI and One Health

- The new areas identified at risk could be used to set targeted & enhanced surveillance and preventive measures by health authorities
- Sharing the code and providing training on the model back to health authorities increases usefulness, trust and impact of the research outputs from large scale projects
- Further works should incorporate data streams from animals and mosquitos – One Health

Horizon2020
European Union Funding
for Research & Innovation

MOOD

Special thanks to Kyla Serres (now ULB), and the co-authors from ULB and ITM, Belgium & ERGO, UK


Surveillance

Integrating indicator-based and event-based surveillance data for risk mapping of West Nile virus, Europe, 2006 to 2021

Check for updates

Like

Kyla Serres[1,2,*], Diana Erazo[1,*], Garance Despréaux[1,*], María F Vincenti-González[1], Wim Van Bortel[3,4], Elena Arsevska[2,**], Simon Dellicour[1,5,**]

Many thanks to the PADI-web team

tetis
TERRITOIRE ENVIRONNEMENT TÉLÉDÉTECTION INFORMATION SPATIALE

ASTRE
Animal · Santé
Territoires · Risques · Écosystèmes

PADI-web contact

Further questions and ideas on collaborations

padi-web@cirad.fr

https://www.padi-web-one-health.org

PADI - WEB

cirad
LA RECHERCHE AGRONOMIQUE POUR LE DÉVELOPPEMENT

ULB UNIVERSITÉ LIBRE DE BRUXELLES

INSTITUTE OF TROPICAL MEDICINE ANTWERP

·E·R·G·O·
Environmental Research Group Oxford