University of Trento
C3A - Center Agriculture Food Environment

MACHINE LEARNING AND STACKED GENERALISATION

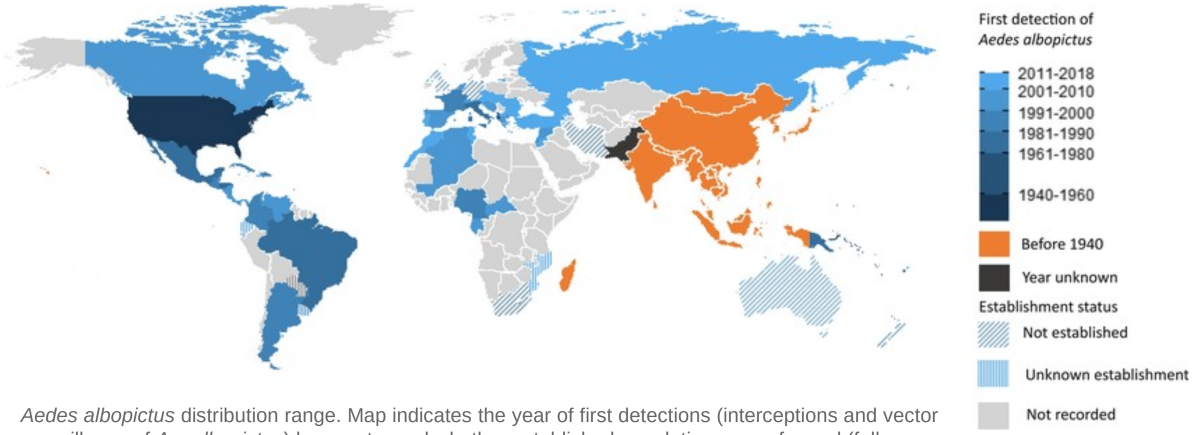
Forecasting the spatio-temporal abundance of *Aedes albopictus*

Marharyta Blaha

September 19, 2024

Introduction to *Ae. albopictus*

- *Aedes albopictus*, also known as the Asian tiger mosquito, is a **highly invasive** species and a **vector for several viruses**, including dengue, Zika, and chikungunya.

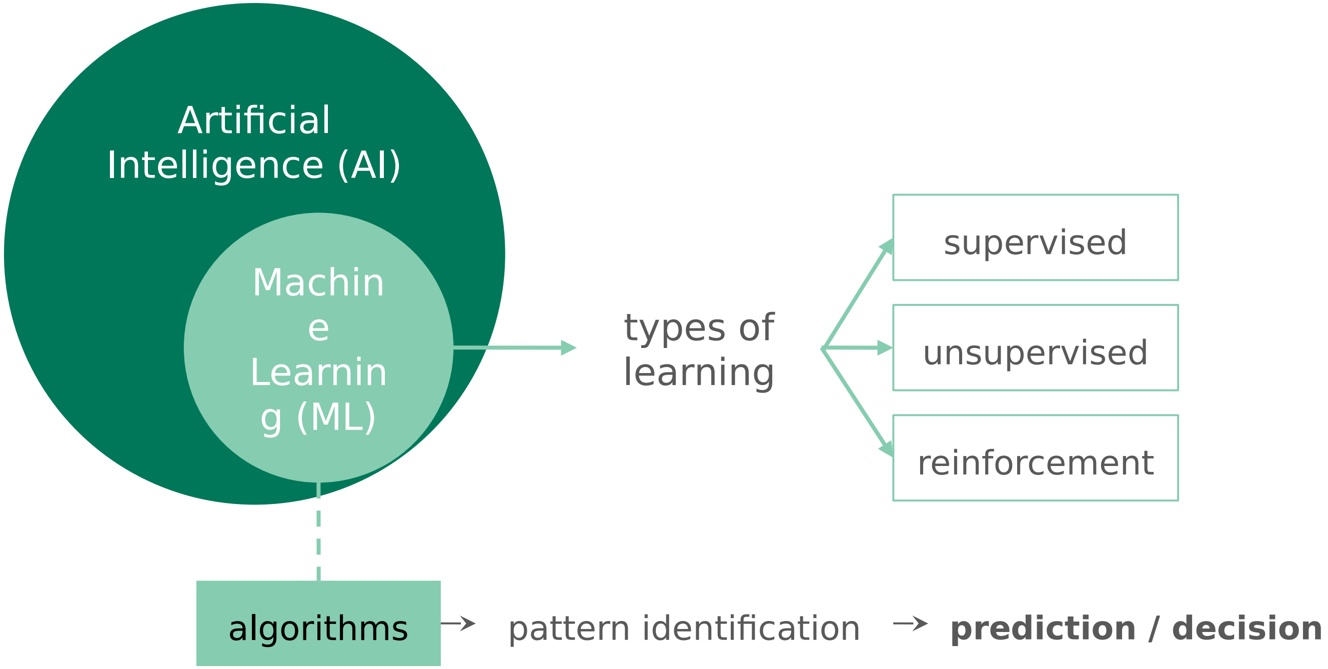


Aedes albopictus distribution range. Map indicates the year of first detections (interceptions and vector surveillance of *Ae. albopictus*) by country and whether established populations were formed (full colour).
Source: https://parasitesandvectors.biomedcentral.com/articles/10.1186/s13071-022-05413-5

Introduction to *Ae. albopictus*

- *Aedes albopictus*, also known as the Asian tiger mosquito, is a **highly invasive** species and a **vector for several viruses**, including dengue, Zika, and chikungunya.
- Understanding its spatio-temporal distribution is crucial for **public health** planning and **vector control** strategies.

A brief introduction to Machine Learning (ML)



Machine learning (ML) in ecology

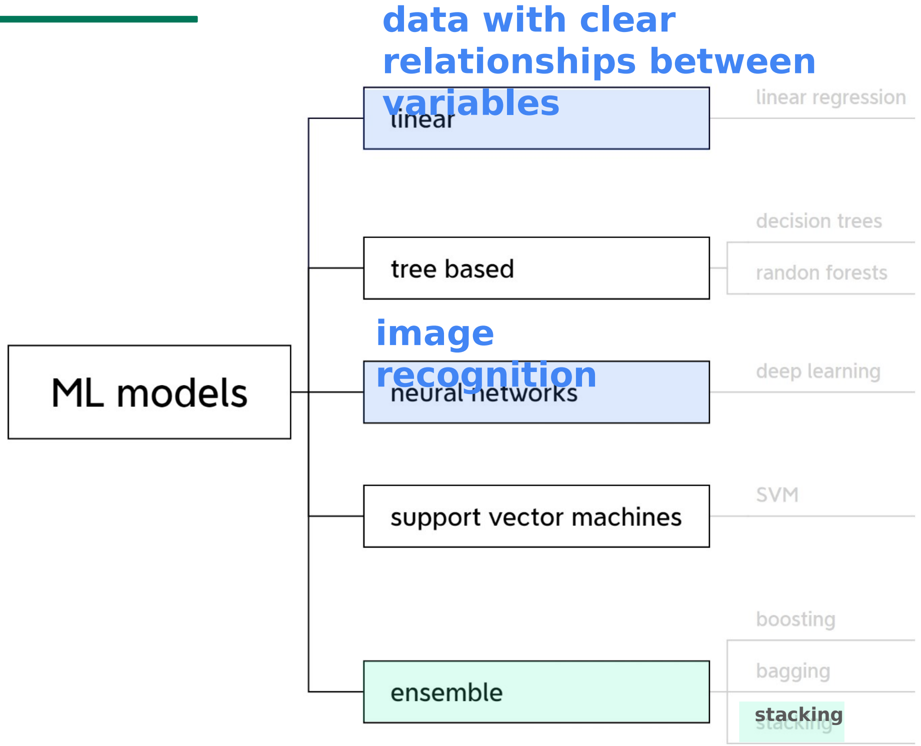
Conventional statistical models

Pros	Cons
Predefined, explicit assumptions and relationships. Easier to interpret.	Struggle with complex, non-linear interactions and large datasets.

Machine Learning

Pros	Cons
Capture of complex patterns. High-dimensional datasets.	More difficult to interpret (e.g. causality)

Which model is “the best”?



combine the strengths of multiple models

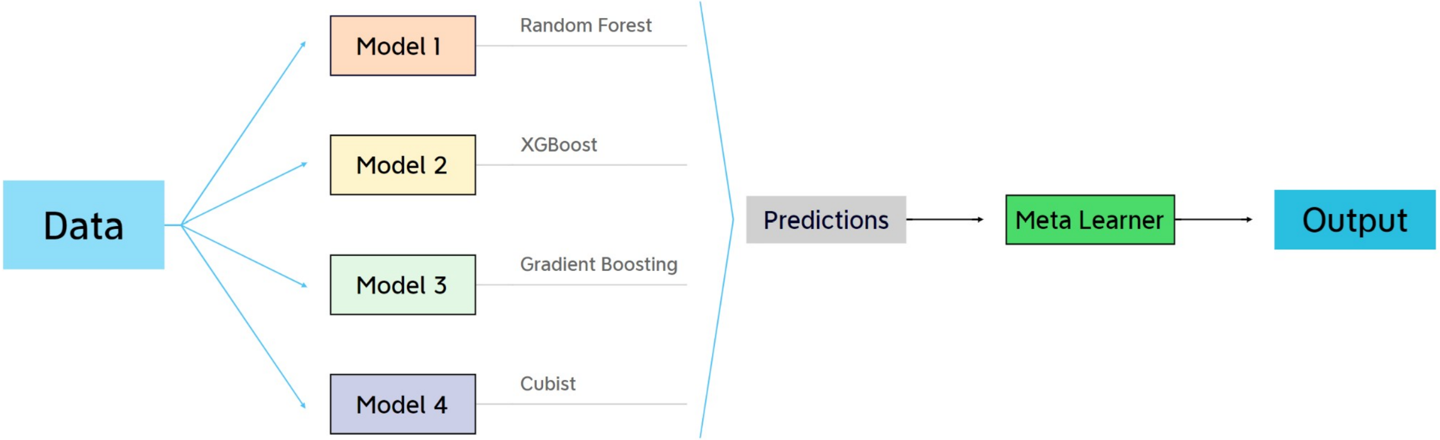
No single model is "the best"; the choice depends on:

1. problem
2. data quality
3. goals

Stacked generalisation

Stacked generalisation (stacking) combines multiple models to improve prediction accuracy.

It works by training a **meta-model** to learn from the predictions of **base models**, effectively reducing overfitting and bias.



What kind of model?

- **Stacked** machine learning models have already been implemented and tested in several studies (e.g. *Ae. albopictus* in Southern Europe);

Forest tree species distribution for Europe 2000–2020: mapping potential and realized distributions using spatiotemporal machine learning

Carmelo Bonannella [1] [2], Tomislav Hengl [2], Johannes Heisig [3], Leandro Parente [2], Marvin N Wright [4] [5], Martin Herold [1] [6], Sytze de Bruin [1]

Affiliations + expand

PMID: 35910765 PMCID: PMC9332400 DOI: 10.7717/peerj.13728

Inferring the seasonal dynamics and abundance of an invasive species using a spatio-temporal stacked machine learning model

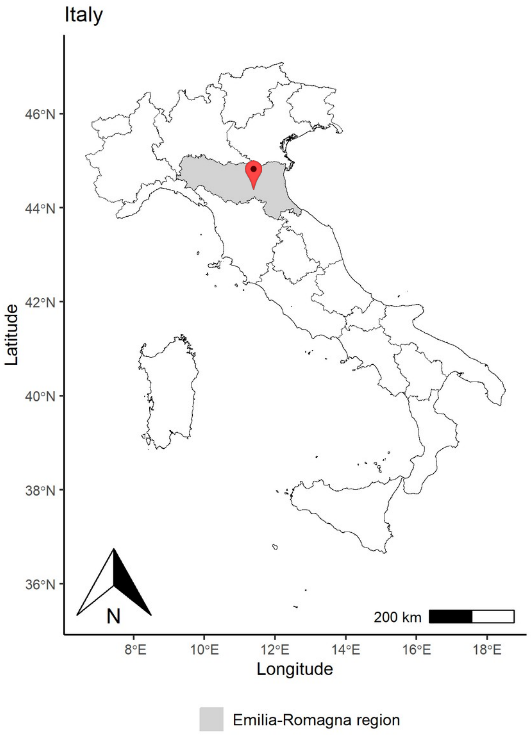
Daniele Da Re[1,2], Giovanni Marini[2,3], Carmelo Bonannella[4,5], Fabrizio Laurini[6], Mattia Manica[3,7], Nikoleta Anicic[8], Alessandro Albieri[9], Paola Angelini[10], Daniele Arnoldi[2], Federica Bertola[11], Beniamino Caputo[12], Claudio De Liberato[13], Alessandra della Torre[12], Eleonora Flacio[8], Alessandra Franceschini[14], Francesco Gradoni[15], Përparim Kadriaj[16], Valeria Lencioni[14], Irene Del Lesto[13], Francesco La Russa[17], Riccardo Paolo Lia[18], Fabrizio Montarsi[15], Domenico Otranto[18], Gregory L'Ambert[19], Annapaola Rizzoli[2,3], Pasquale Rombolà[13], Federico Romiti[13], Gionata Stancher[11], Alessandra Torina[17], Enkelejda Velo[16], Chiara Virgillito[12], Fabiana Zandonai[11], Roberto Rosà[1,2]

What kind of model?

- **Stacked** machine learning models have already been implemented and tested in several studies (e.g. *Ae. albopictus* in Southern Europe);
- However, in this study we focus solely on the Emilia-Romagna region;
- Specifically, the aim of this project is
  - to conduct a sensitivity analysis to discern the quantity of data required for reliable estimates of egg distribution and abundance.
  - use the model selected through the sensitivity analysis to forecast *Ae. albopictus* egg abundance over medium (seasonal) and short (weekly) periods.

Case Study: *Aedes albopictus* in Emilia-Romagna

- What data?
- What model(s)?
- ML workflow
- Results: metrics and maps



What data?

<u>Response variable</u>

- Ovitrap observations

<u>Predictors</u>

- ARPAE historical average weekly temperature (median)*
- ARPAE historical weekly precipitation (sum)*
- Photoperiod*
- Fourier harmonics (seasonal and interannual)
- Urbanization Index (ESA CCI landcover)

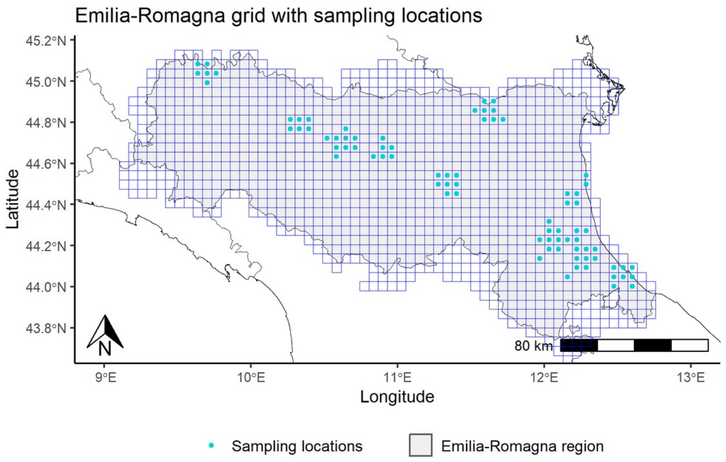
* lagged weekly data (2 and 3), as current distribution of the target variable depends on past values

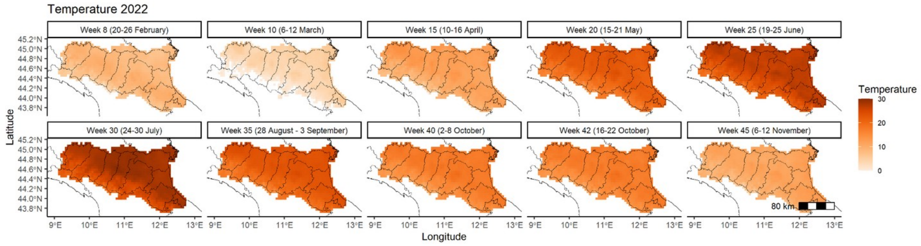
<u>Spatial resolution</u>

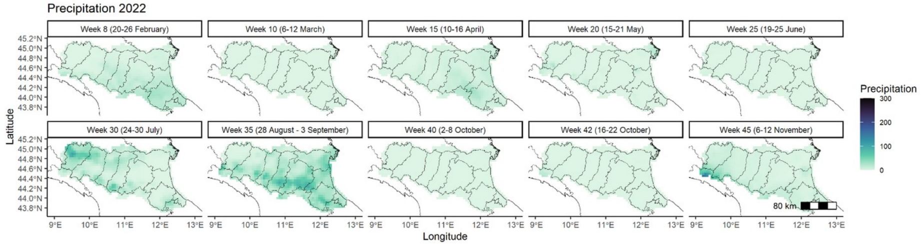
- 5 km grid of Emilia-Romagna

<u>Temporal resolution</u>

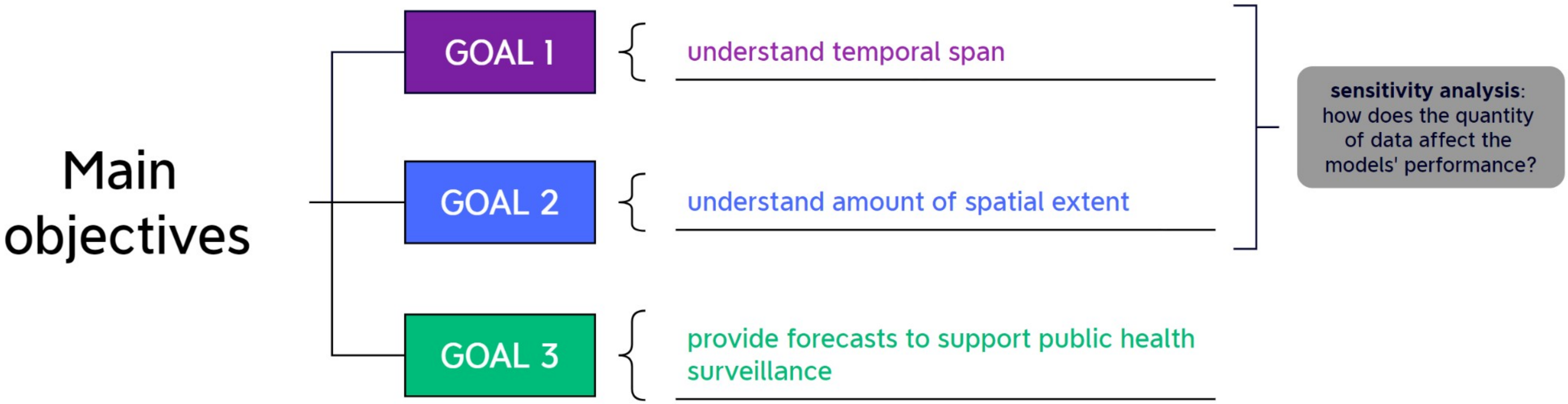
- 1 week



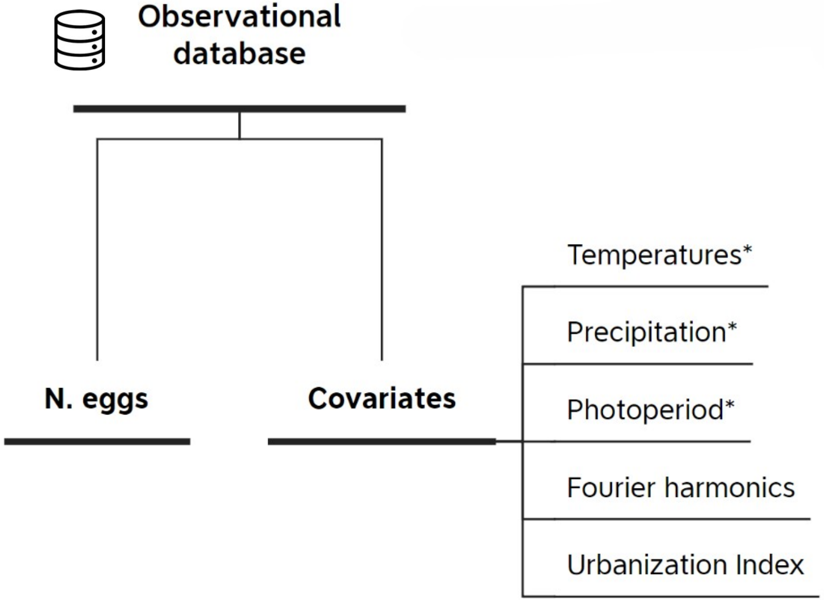




To sum up:

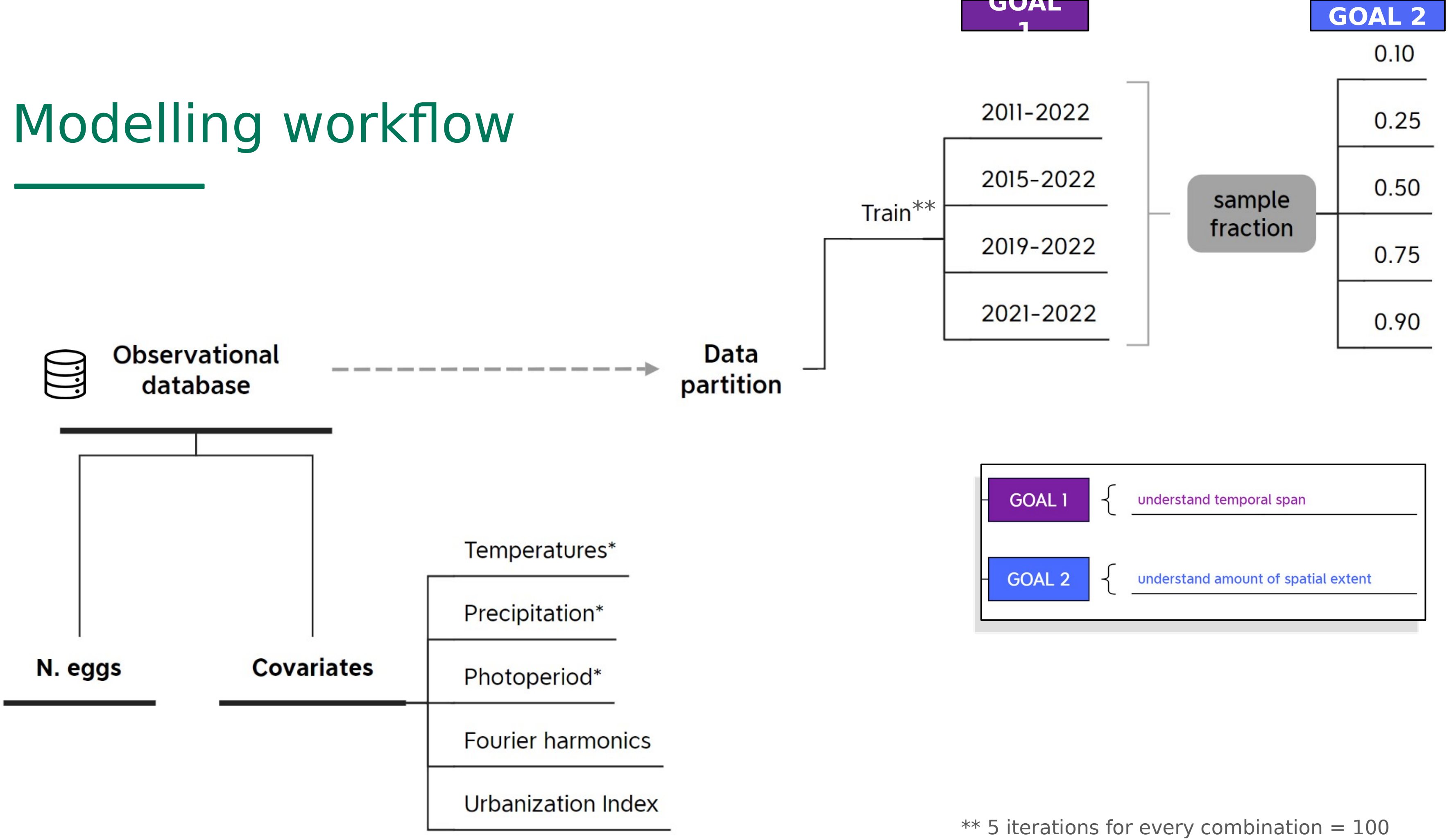

Main objectives
GOAL 1
understand temporal span
GOAL 2
understand amount of spatial extent
GOAL 3
provide forecasts to support public health surveillance
sensitivity analysis: how does the quantity of data affect the models' performance?

Modelling workflow



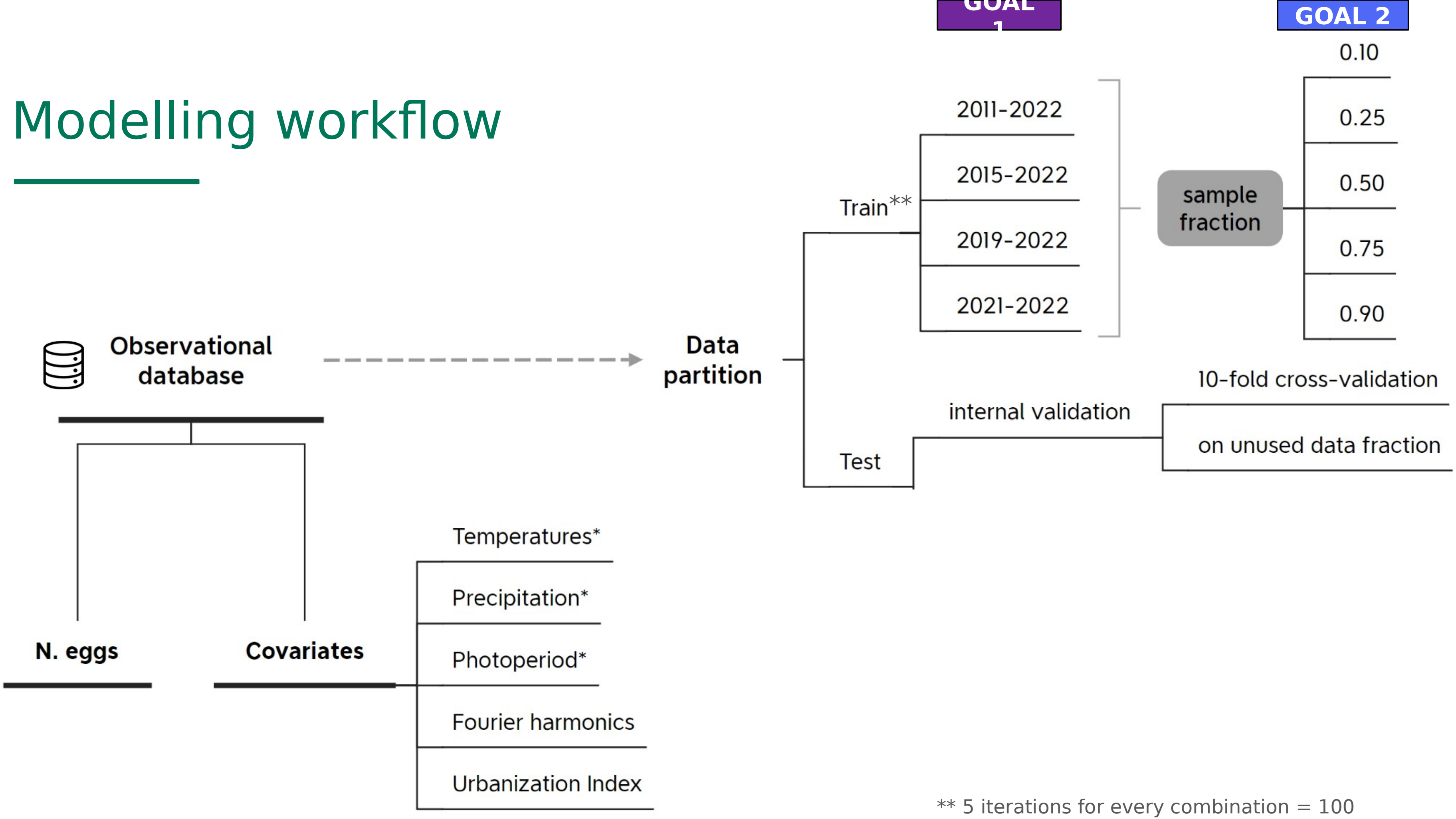
* lagged weekly data (2 and 3), as current distribution of the target variable depends on past values

Modelling workflow



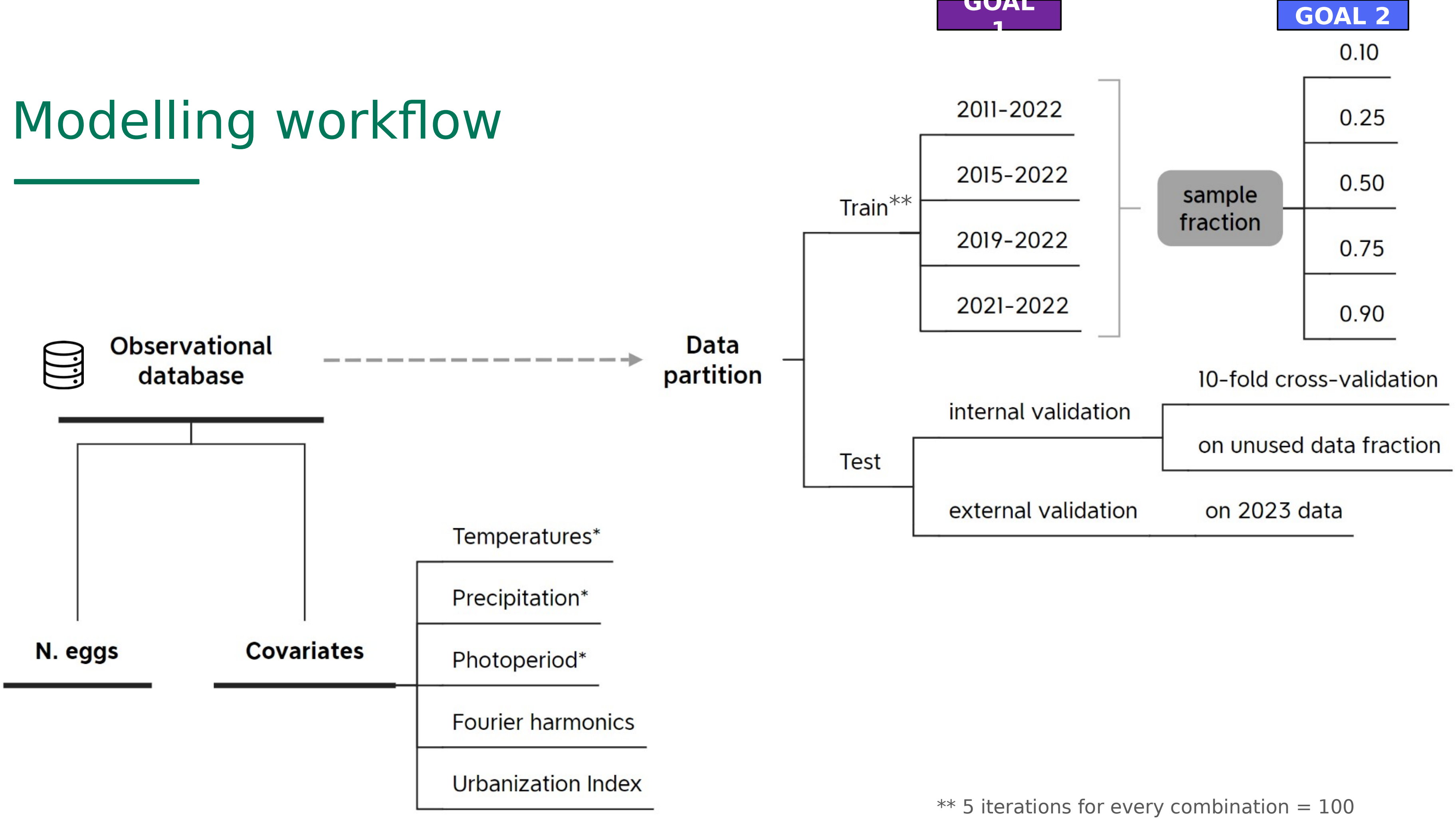
** 5 iterations for every combination = 100 models

Modelling workflow



** 5 iterations for every combination = 100 models

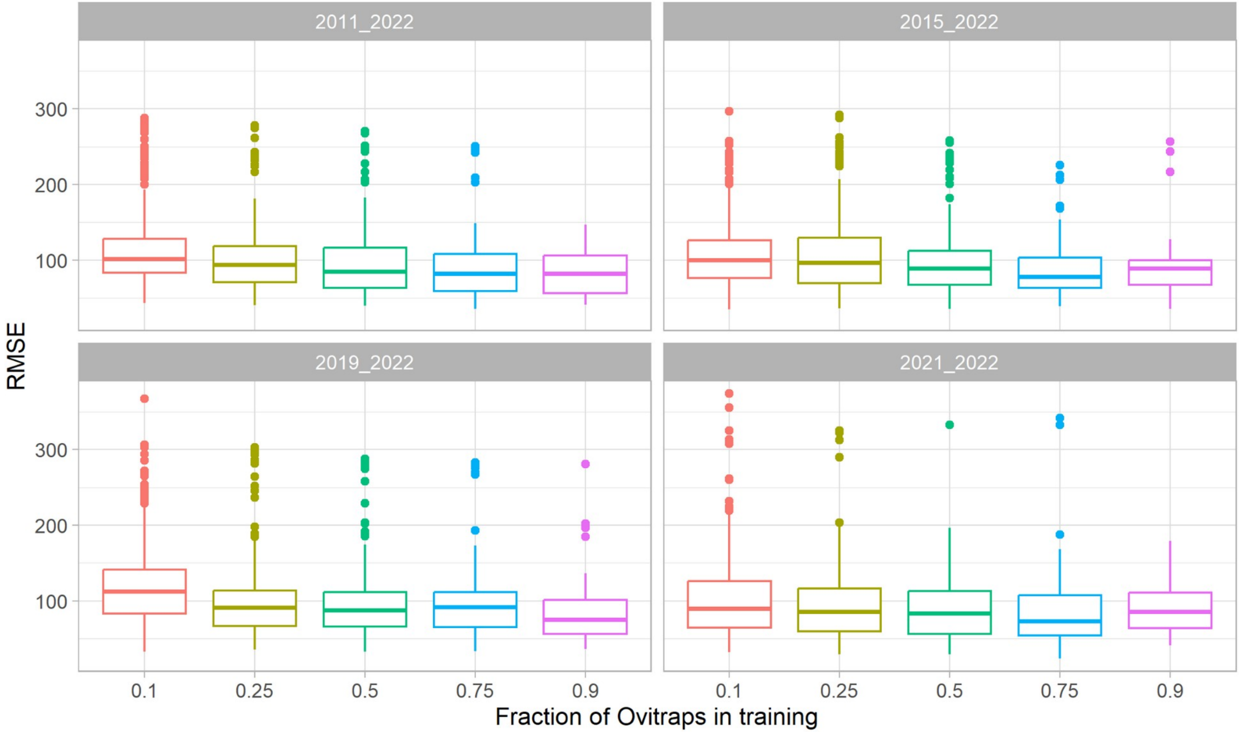
Modelling workflow



** 5 iterations for every combination = 100 models

Results: RMSE for fraction of ovitraps

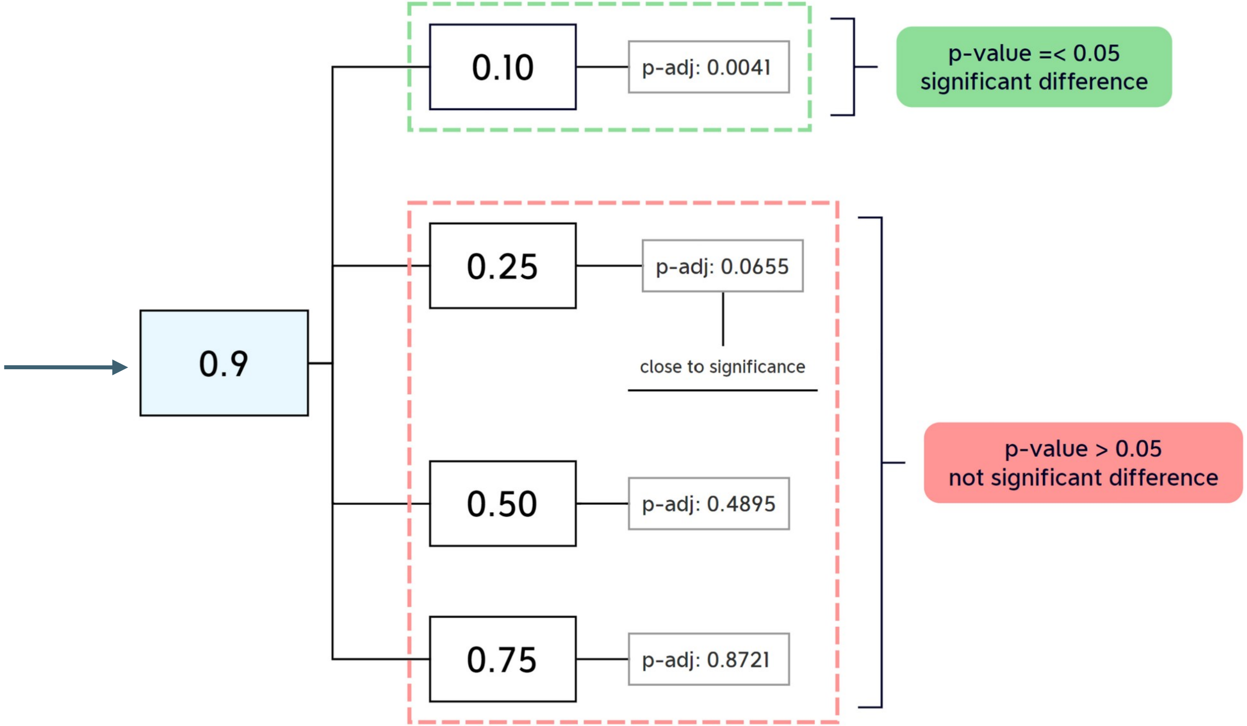
- Comparable scale of error;
- **ANOVA**: both training years and fraction have a statistically significant effect ($p < 0.001$);
- Clear trend of decreasing RMSE as the fraction increases;
- Lowest RMSE: 0.9 fraction (mean RMSE: 89.35).



Results: RMSE for fraction of ovitraps

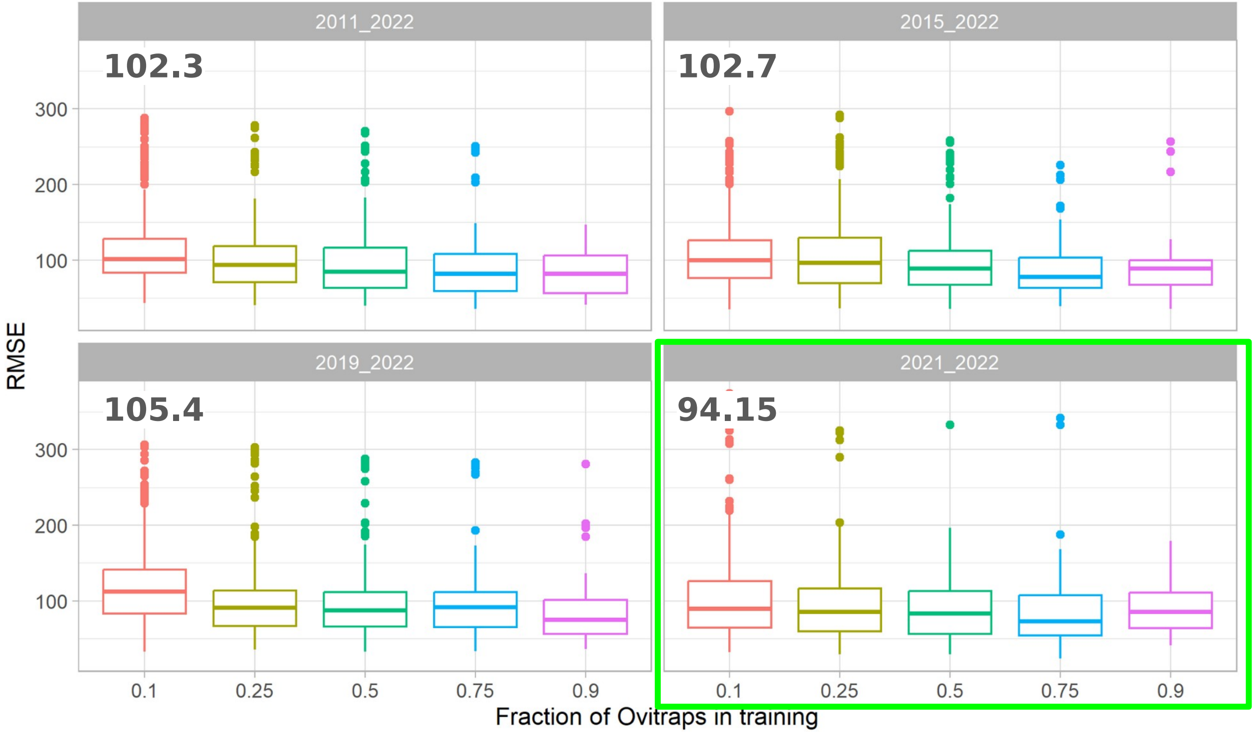
The **Tukey HSD:** 0.9 fraction significantly outperforms lower fractions.

Q: How significant is the difference in error between 0.9 and other fractions?



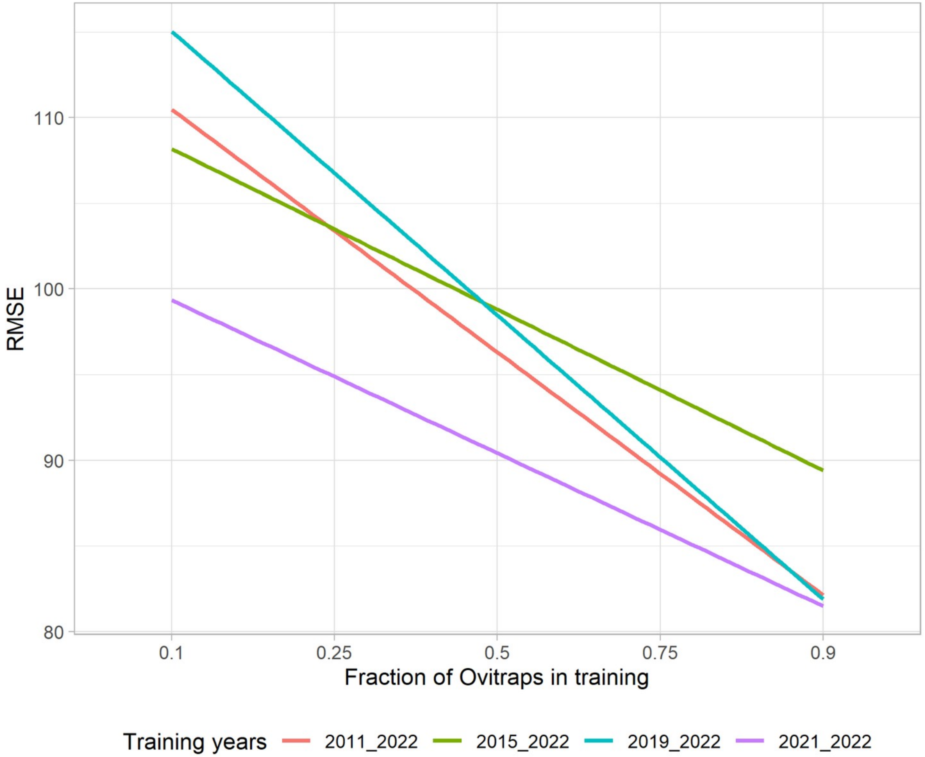
Results: RMSE for years of training data

The lowest mean RMSE is observed for the **2021-2022 training period**.



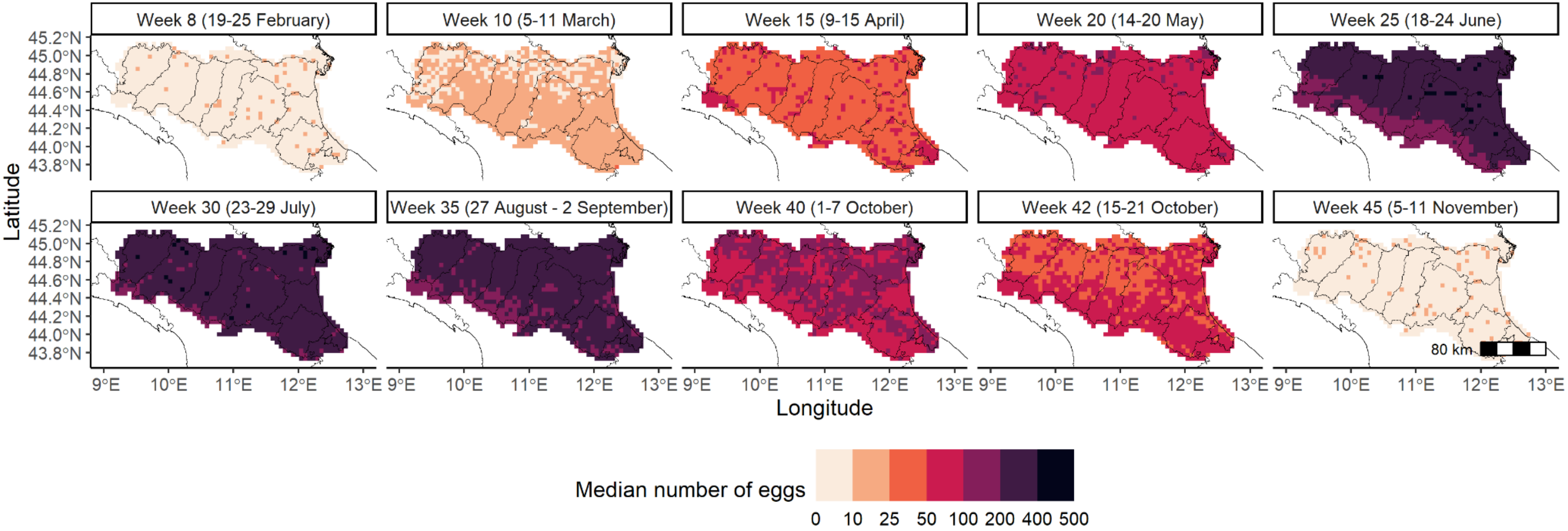
Results: RMSE for years of training data

Is a model trained on 2 years of data better than the one trained on 12 years?



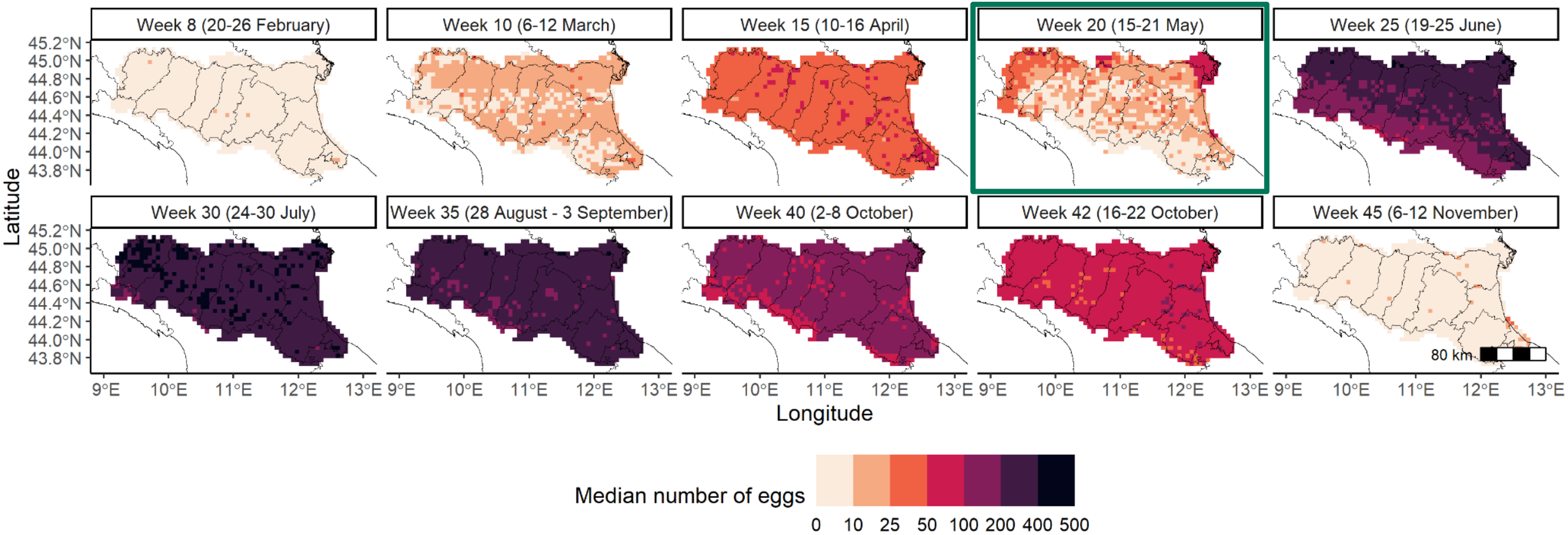
Medium term predictions on 2023

TRAINING ON YEARS 2021-2022 AND FRACTION 0.9



Medium term predictions on 2023

TRAINING ON YEARS 2011-2022 AND FRACTION 0.9



Medium term predictions on 2023


Mappa del sito | Accessibilità | Contatti | Italiano | English

ISPRA

Istituto Superiore per la Protezione e la Ricerca Ambientale

Cerca

Home | Istituto | Attività | Servizi | Dati e Indicatori | Pubblicazioni | News | Eventi ISPRA | Sala stampa | URP

Home / Archivio / News / Notizie ispra / Anno 2023 / Maggio / Alluvione in Emilia-Romagna: piogge record, fiumi e corsi d'acqua esondati

Alluvione in Emilia-Romagna: piogge record, fiumi e corsi d'acqua esondati

ISPRA, in costante contatto con le Agenzie coinvolte del SNPA, partecipa come Centro di Competenza ai lavori del Comitato operativo di protezione civile, fornendo supporto tecnico- scientifico. In particolare, ultimata la prima fase del soccorso tecnico ancora in corso, i lavori proseguiranno con la pianificazione delle attività di gestione dei fanghi e dei rifiuti riversati sulle vie di comunicazione a seguito delle esondazioni e con eventuali sopralluoghi nelle aree maggiormente colpite. Il Presidente ISPRA e SNPA Stefano Laporta: "Solidarietà alla popolazione dell'Emilia Romagna in questo momento di grave difficoltà.

Foto: ANSA

ISPRA e SNPA assicurano il massimo supporto tecnico scientifico e la massima disponibilità a fornire al DPC e alle Amministrazioni locali tutte le informazioni in loro possesso in materia di dissesto idrogeologico, consapevoli che la conoscenza e l'informazione possono rendere i territori e chi vi abita maggiormente resilienti a tragedie di questa natura". Maltempo Emilia-Romagna: proseguono le attività di soccorso alle popolazioni colpite

Il territorio dell'Emilia-Romagna è stato interessato da due eventi in sequenza in meno di venti giorni con precipitazione cumulata mensile che ha superato i 450 millimetri in varie località.

L'evento in corso dalla mezzanotte del 15 maggio al 17 maggio ha causato l'esondazione di 21 fiumi e allagamenti diffusi in 37 comuni. Nelle ultime 48 ore si sono registrati picchi di 300 millimetri sui bacini del crinale e collina forlivese. Sulla stessa area, sulle colline e montagna ravvenati e sul settore orientale del bolognese sono in media caduti tra i 150 e i 200 millimetri. Sulla pianura cesenate forlivese fino a 150 millimetri[2].

Complessivamente risultano attive almeno un migliaio di frane, di cui circa 300 più significative concentrate in 54 comuni.

Piattaforma nazionale IdroGEO – Pericolosità e indicatori di rischio su Regione Emilia-Romagna


L'Espresso

Abbonati

EMERGENZA TERRITORIO

Alluvione Emilia-Romagna, l'allarme inascoltato degli esperti: «Metà regione è a rischio»

di Paolo Biondani 19 maggio 2023

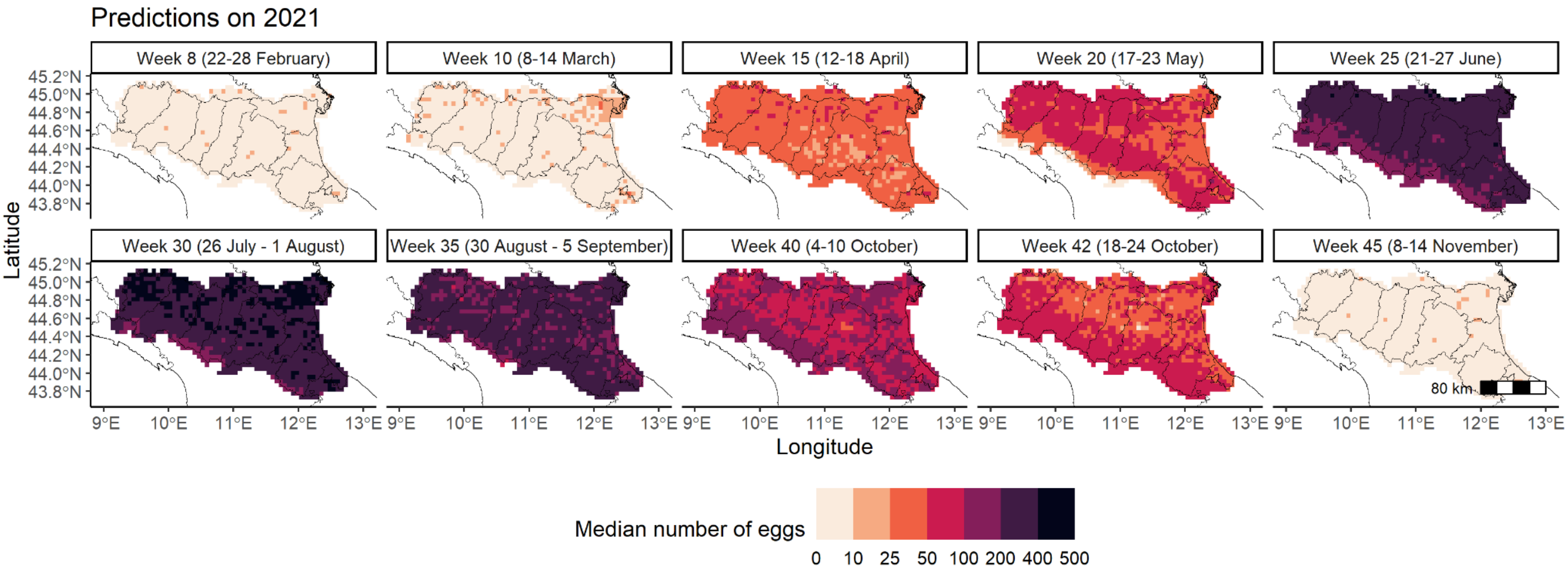
Schedato da anni come "allagabile" il 45 per cento del territorio regionale, dove vivono tre abitanti su cinque. Ma la cementificazione continua, come in tutta Italia. E il cambiamento climatico aumenta i rischi di siccità con successive piogge "cicloniche"

Date start	2 May 2023 (I fase)
	15 May 2023 (II fase)

Date end	3 May 2023 (I fase)
	17 May 2023 (II fase)

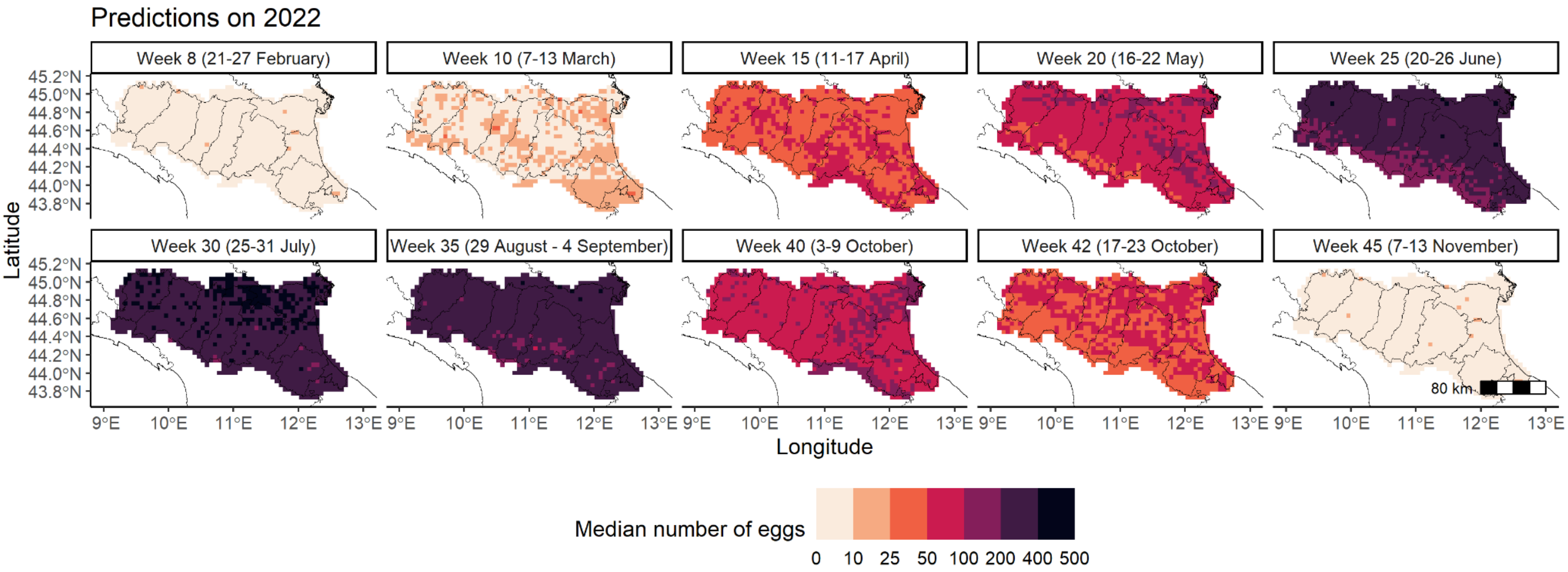
Spatial predictions on 2021

TRAINING ON YEARS 2011-2022 AND FRACTION 0.9



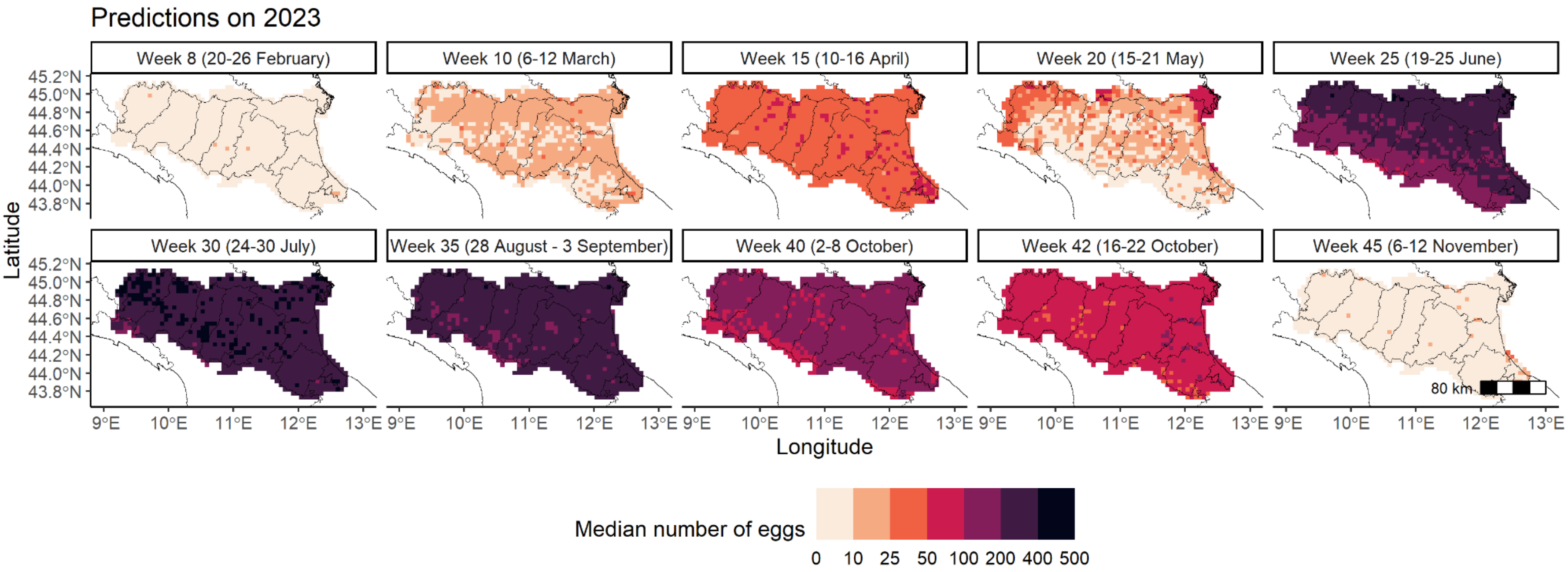
Spatial predictions on 2022

TRAINING ON YEARS 2011-2022 AND FRACTION 0.9



Spatial predictions on 2023

TRAINING ON YEARS 2011-2022 AND FRACTION 0.9



Forecasting for 2024

JUNE - JULY - AUGUST - *SEPTEMBER*

TRAINING ON YEARS 2011-2022 AND FRACTION 0.9

TEMPERATURE AND PRECIPITATION ESTIMATES

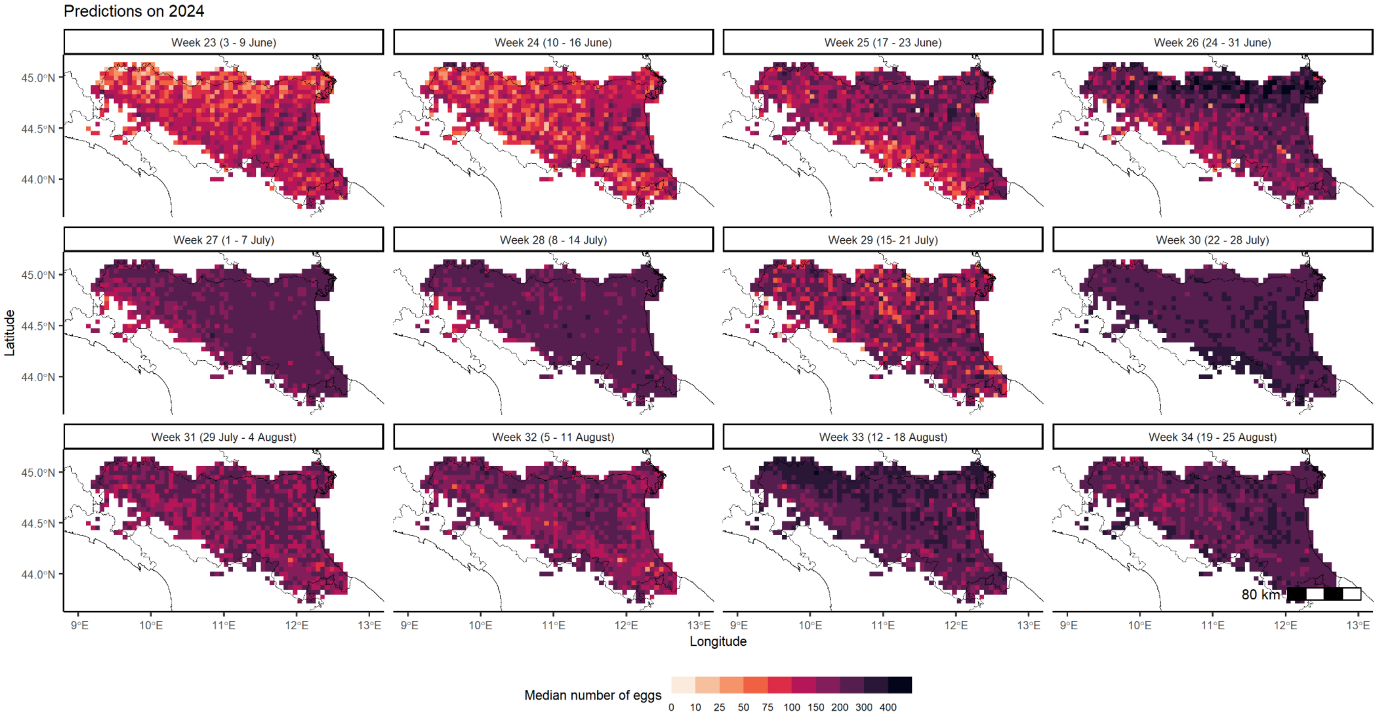
+ lags 2 and 3
+ fourier harmonics
+ urbanisation

AE. ALBOPICTUS EGG ABUNDANCE FOR THE NEXT 3-4 WEEKS

Forecasting for 2024

JUNE - JULY - AUGUST

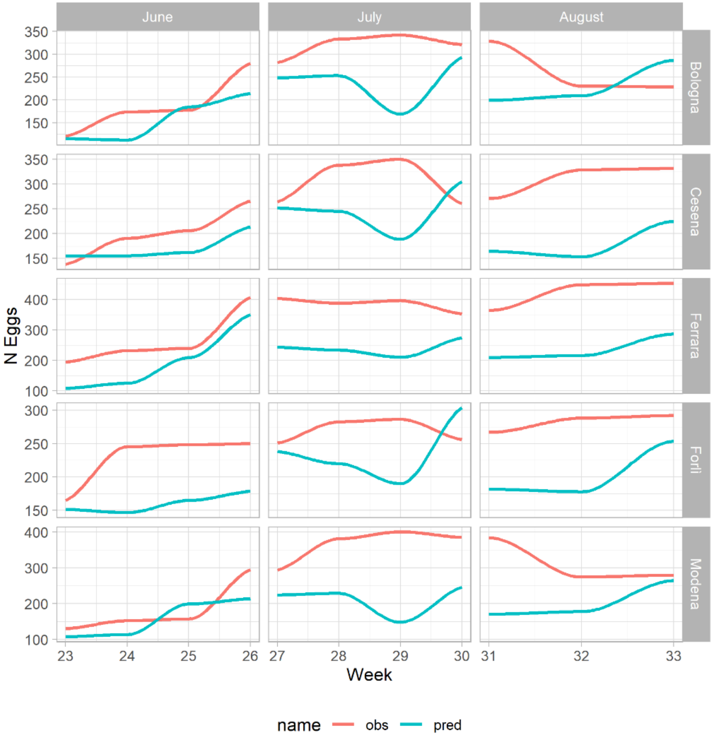
GOAL 3 — provide forecasts to support public health surveillance

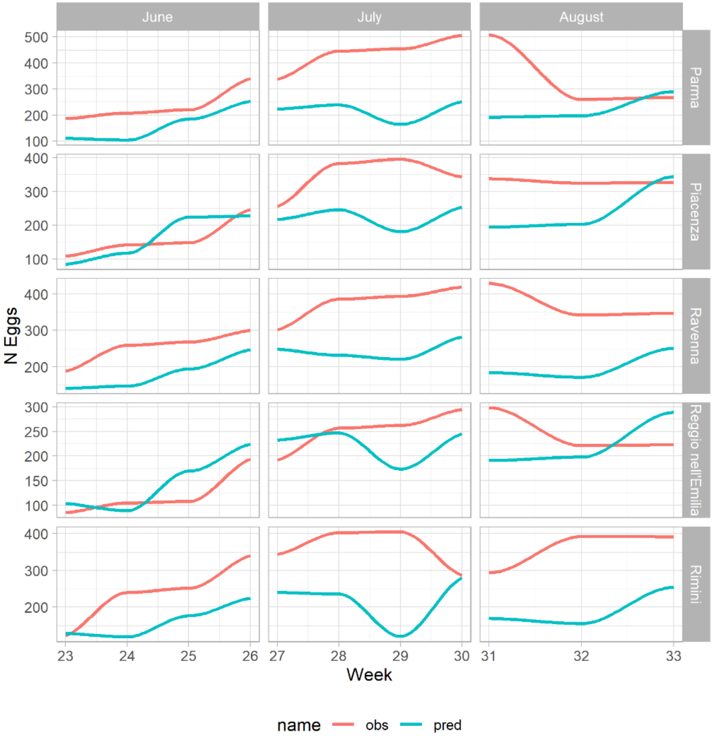


Forecasting for 2024

JUNE - JULY - AUGUST

GOAL 3 — provide forecasts to support public health surveillance





Conclusion

- **ML** show promise in forecasting the spatio-temporal abundance of *Aedes albopictus*;
- **Data driven approaches** allow to save time and expenses;
- Model performance improves with **more data**;
- **Trade-offs** between model complexity, data quantity, and computational efficiency;
- **Stacked generalisation** balances model bias and variance - better generalisation.

Contributors:

Da Re Daniele
Fondazione Edmund Mach

Rosà Roberto
Fondazione Edmund Mach

Marini Giovanni
Fondazione Edmund Mach

Angelini Paola
Emilia Romagna region

arpae emilia-romagna

Regional Environmental
Protection Agency

Bonannella Carmelo
Wageningen University & Research

Albieri Alessandro
Centro Agricoltura
Ambiente "G.Nicoli"

Laurini Fabrizio
University of Parma

Thank you for your attention!

Q & A