

Regional Flood Impact Prediction

Vasileios Vatellis

Supervisors: I. Papoutsis, S. Karozis, I. Klampanos, A. Troumpoukis

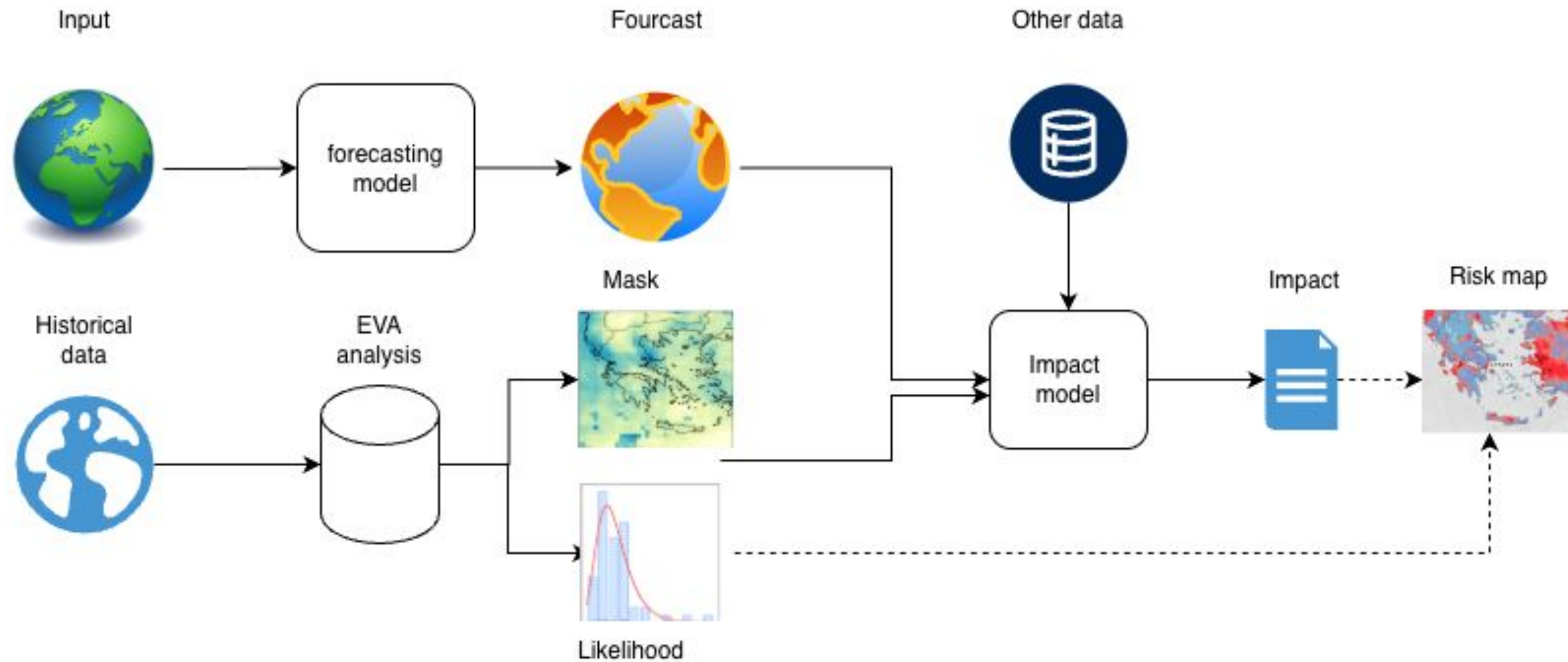
What we need, in order to do Impact prediction?



Forecasting of
Extremes



Impact effects



1 Flag extremes with EVA

2 Forecast the extreme

3 Predict the impact

- NWP remains the reference
- AI models: faster, cheaper (after training)
- Extremes still hard to forecast
- Precipitation: sparse, non-Gaussian
- **Goal: forecast the impact**

PART 1

Forecasting of Extreme Precipitation

- Defining extremes with EVA
- AI weather forecasting models
- Model comparison over Greece

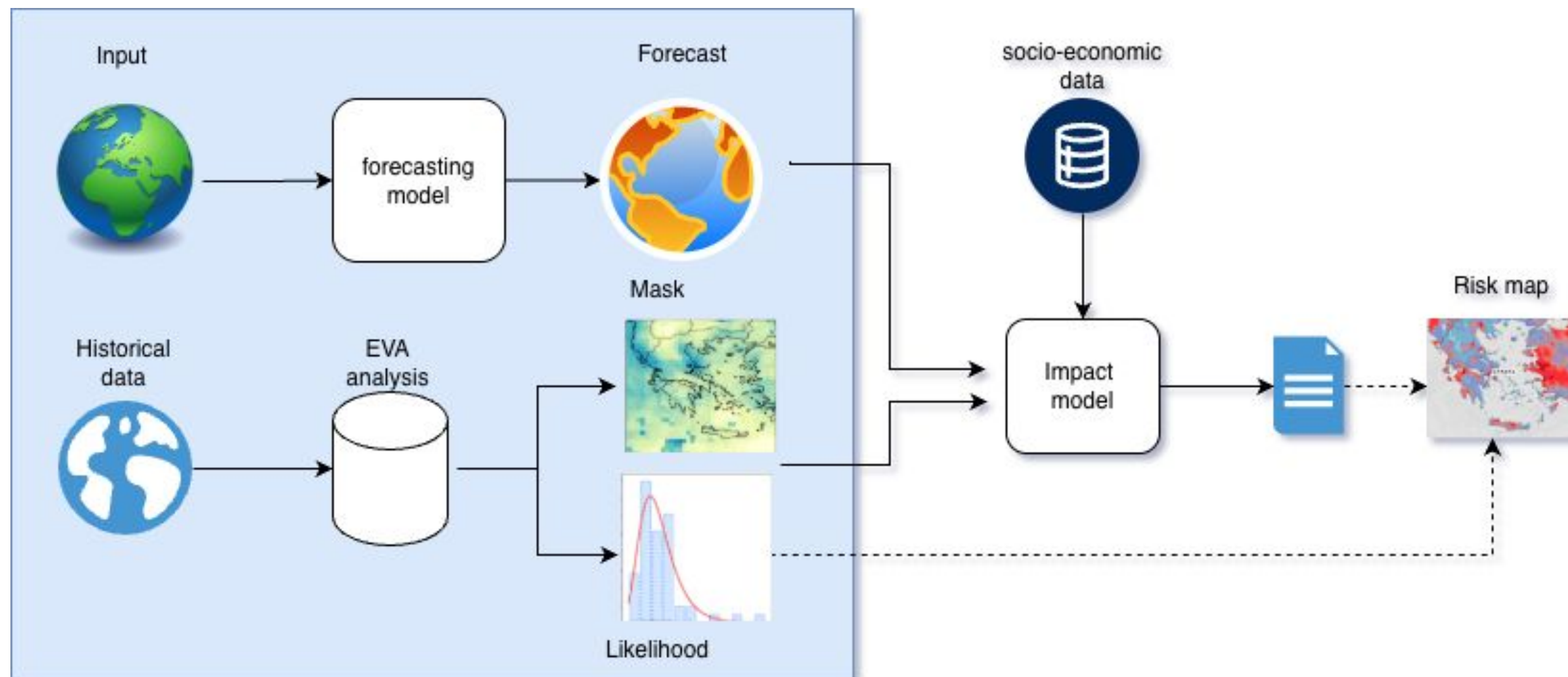
PART 2

Regional Flood Impact Prediction

- Impact-oriented flood dataset
- Two-stage impact classifier
- Greece as spatial holdout

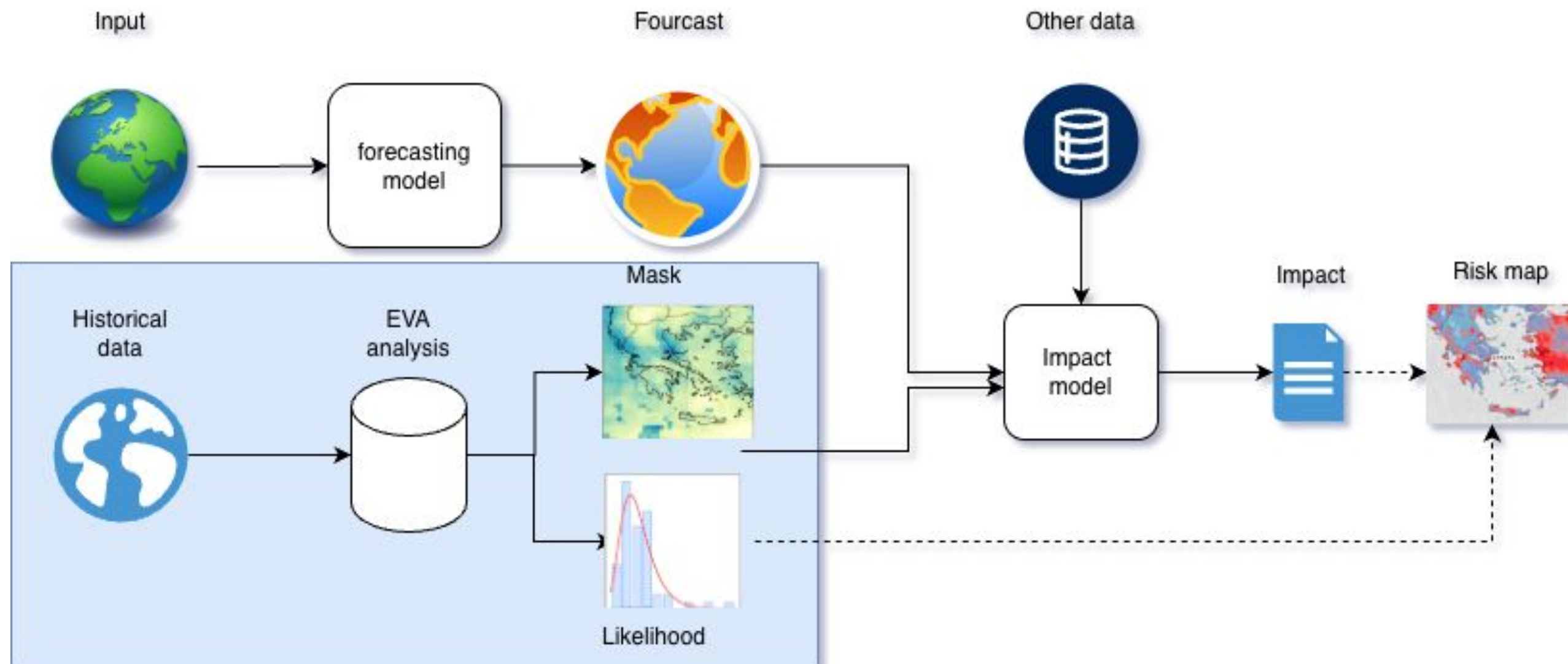
PART 1

Forecasting of Extreme Precipitation



PART 1.A

Extreme Value Analysis (EVA)



Defining Extremes

ERA5

Reanalysis data from 1970-2023

Thresholds

EVA analysis from ERA5 data

Scale

Estimated per grid point

Extreme

At list 10 extreme pixels



Masks

Extreme precipitation dataset

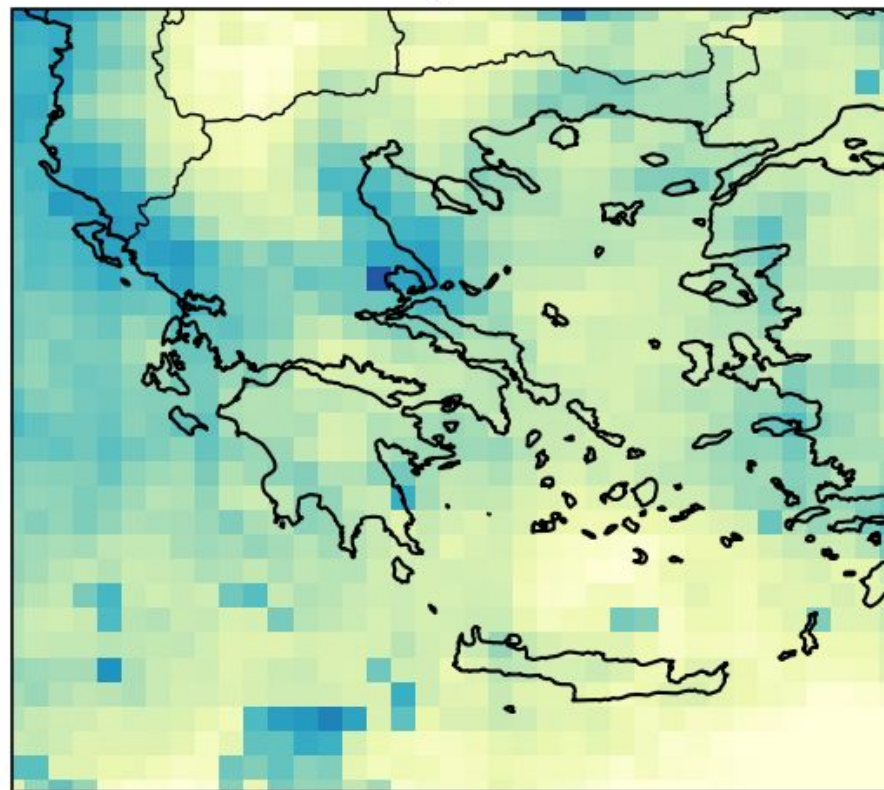
Extreme Precipitation Masks

Variable: total precipitation (tp)

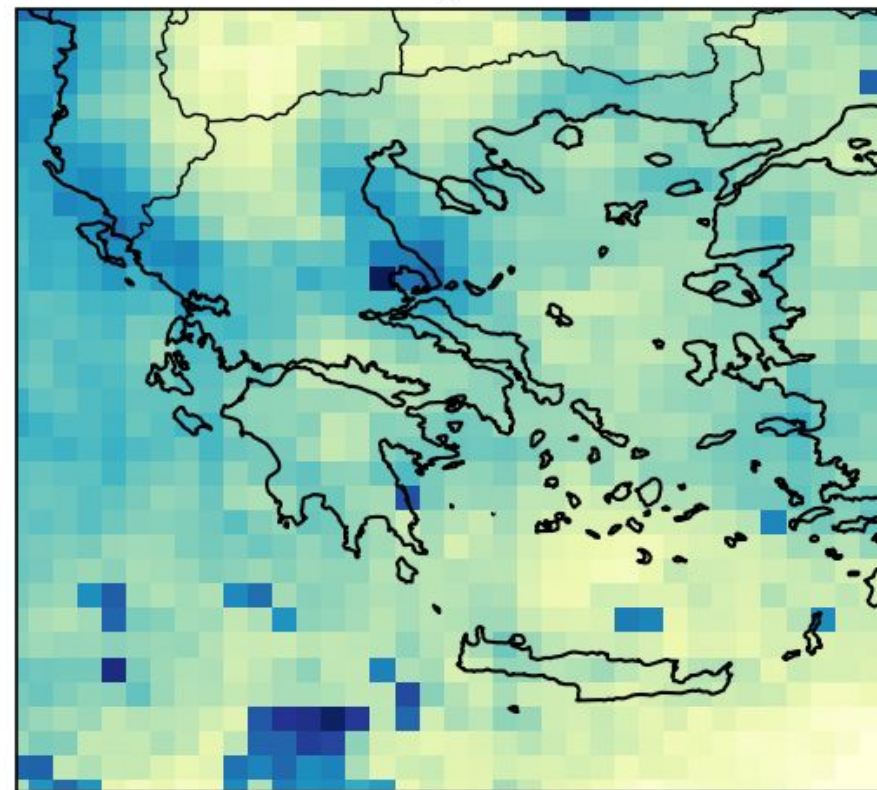
Pixel-level thresholds

10-, 20-, 50-year return levels

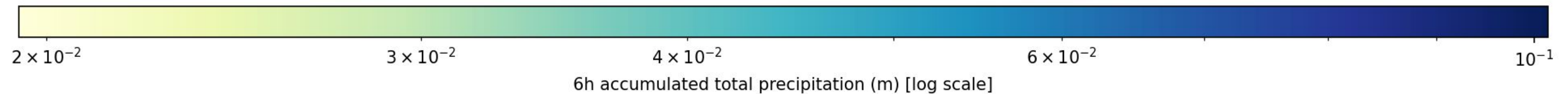
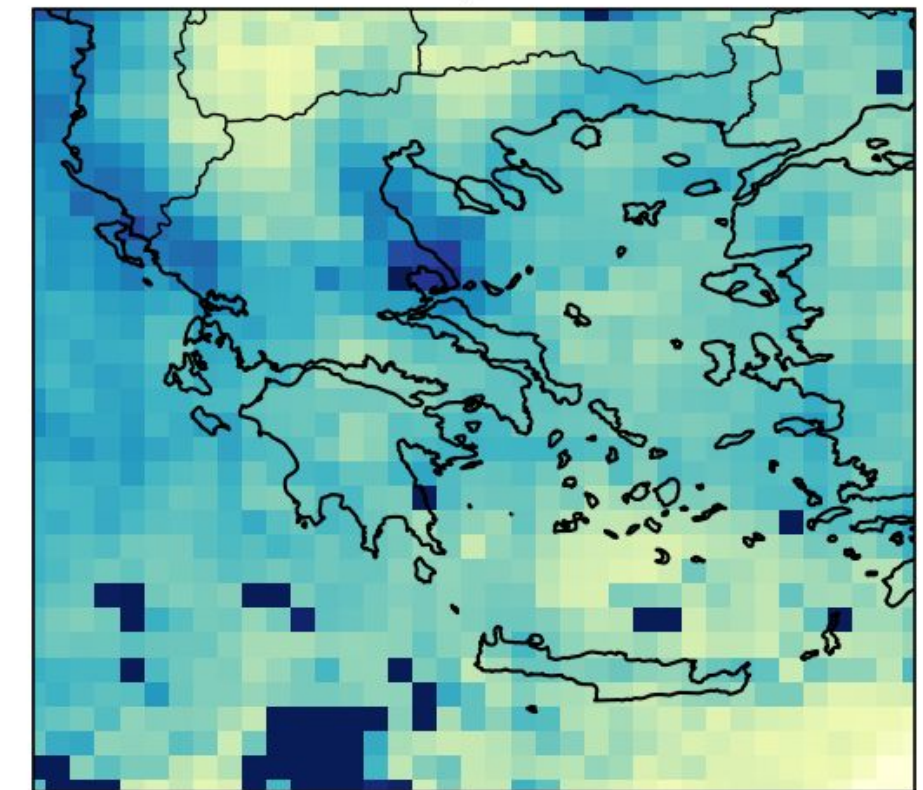
10-year



20-year

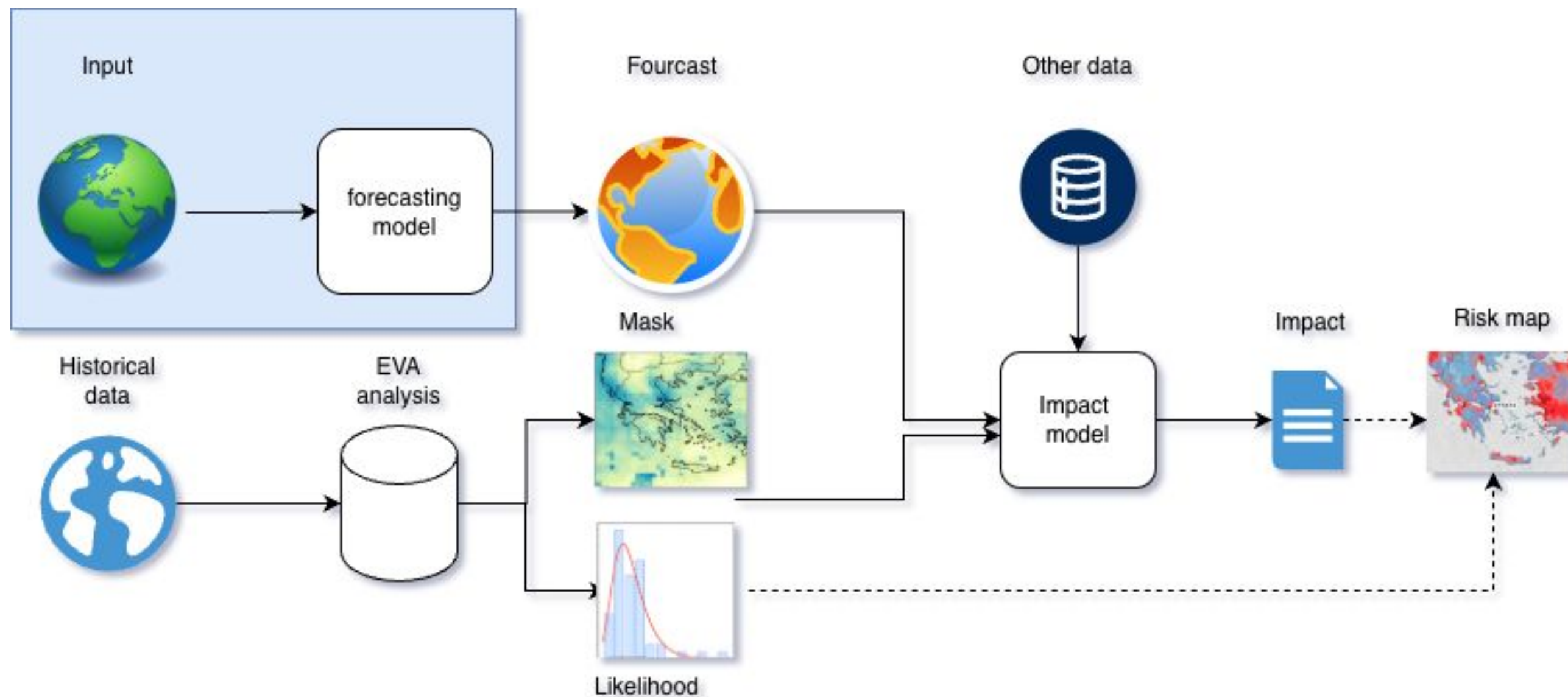


50-year



PART 1.B

Data driven Global Weather Forecasting Models



Data-Driven Forecasting Models

FourCastNet

Fourier neural operator

Pathak et al., 2022

Pangu-Weather

3D Earth-specific transformer

Bi et al., 2023

GraphCast

Graph neural network

Lam et al., 2023

GenCast

Diffusion-based ensemble

Price et al., 2025

FuXi

Cascaded U-Transformer

Chen et al., 2023

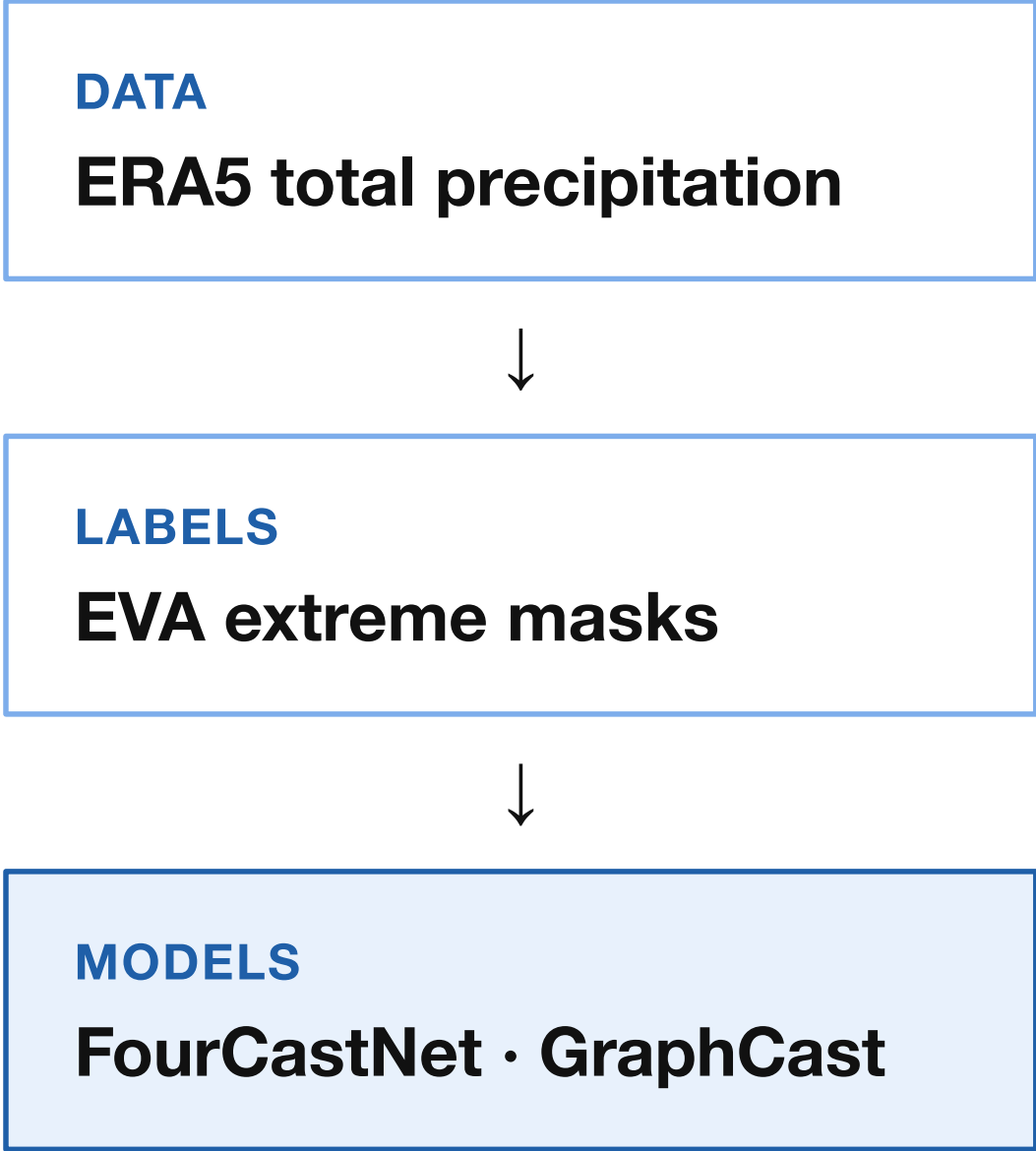
Aurora

Earth-system foundation model

Bodnar et al., 2025

 Total precipitation output: FourCastNet, GraphCast, GenCast

Experimental Setup



Class \ Lead time	1 d	2 d	4 d	6 d	10 d
Non-extreme					
Below threshold					
10-year Return level					
20-year Return level					
50-year Return level					

4 classes

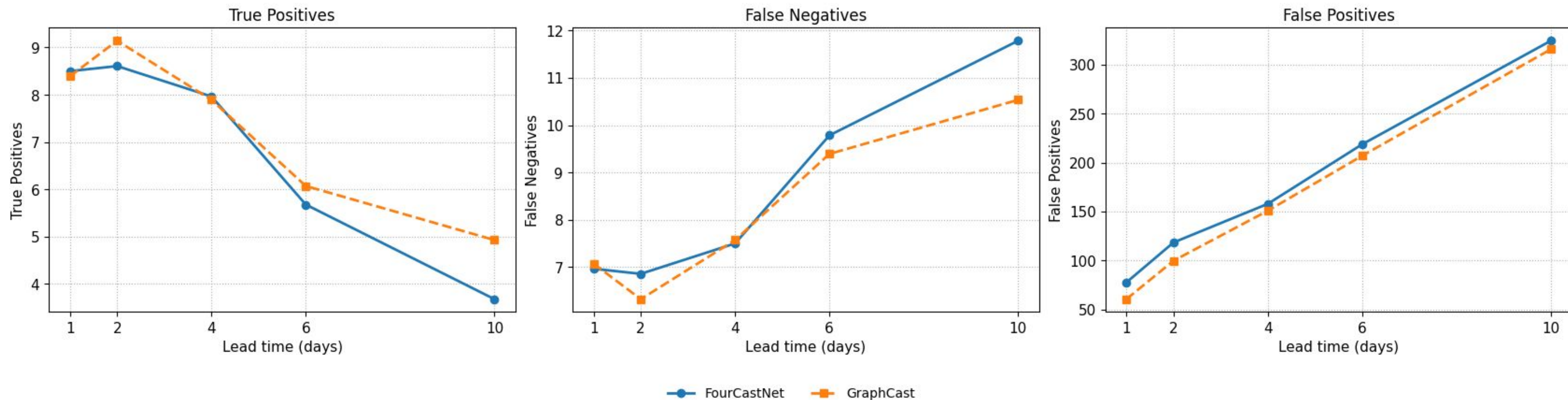
5 lead times

2 models

10 forecasts per event

Data-Driven Forecasting Models

Events with 50-year return levels.



$$score_{rbs} = |SSIM| \times |RMSE| \times |Bias|$$

$$\Delta score = score_{rbs}^{FC} - score_{rbs}^{GC}$$

SSIM

Structural Similarity Index Measure,
comparing their structural information

RMSE

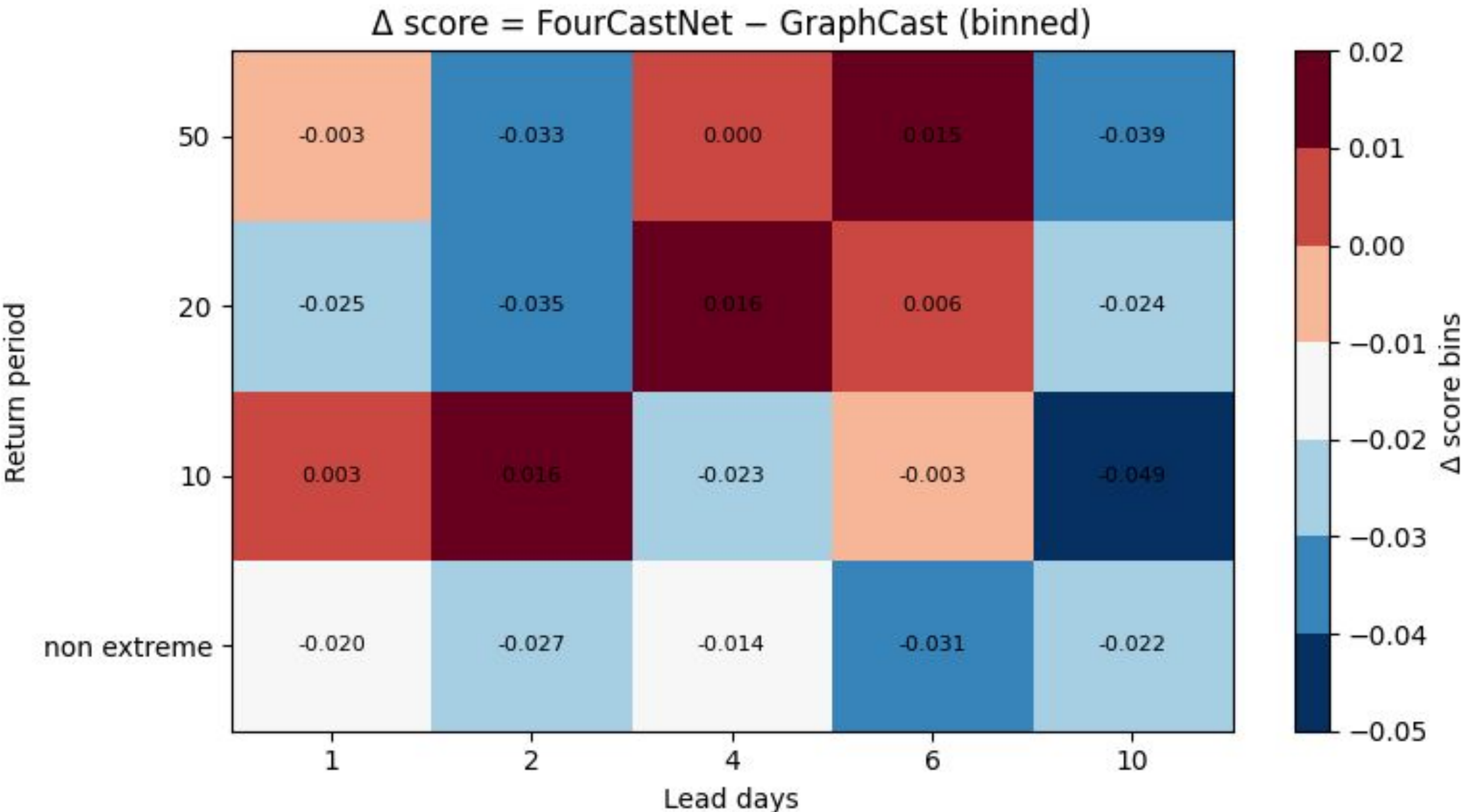
Root mean square error , Magnitude
error

Bias

Systematic offset

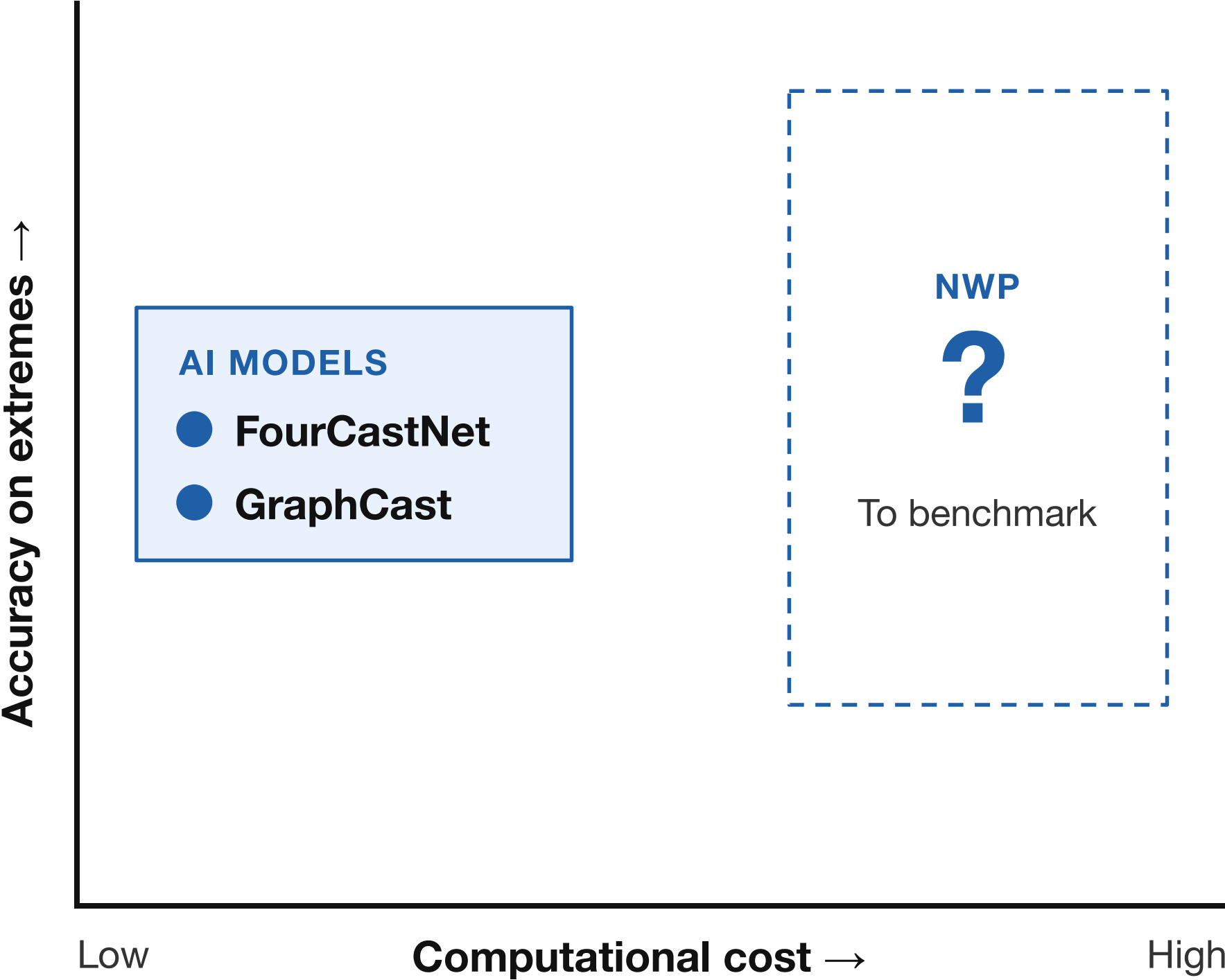
All terms rescaled to [0, 1] · FC: FourCastNet · GC: GraphCast

FourCastNet vs GraphCast



- Use case: Greece
- Lead times: 1–10 days
- Return periods: 10–50 years
- Extremes: Δ mostly negative
- Non-extremes: Δ positive

Positive values indicate higher skill for FourCastNet, negative values for GraphCast



Schematic · not to scale

DONE

First AI benchmark

FourCastNet vs GraphCast, Greece

MISSING

Station observations

EMY data not obtained

NEXT

Compare against NWP

Cost vs accuracy trade-off

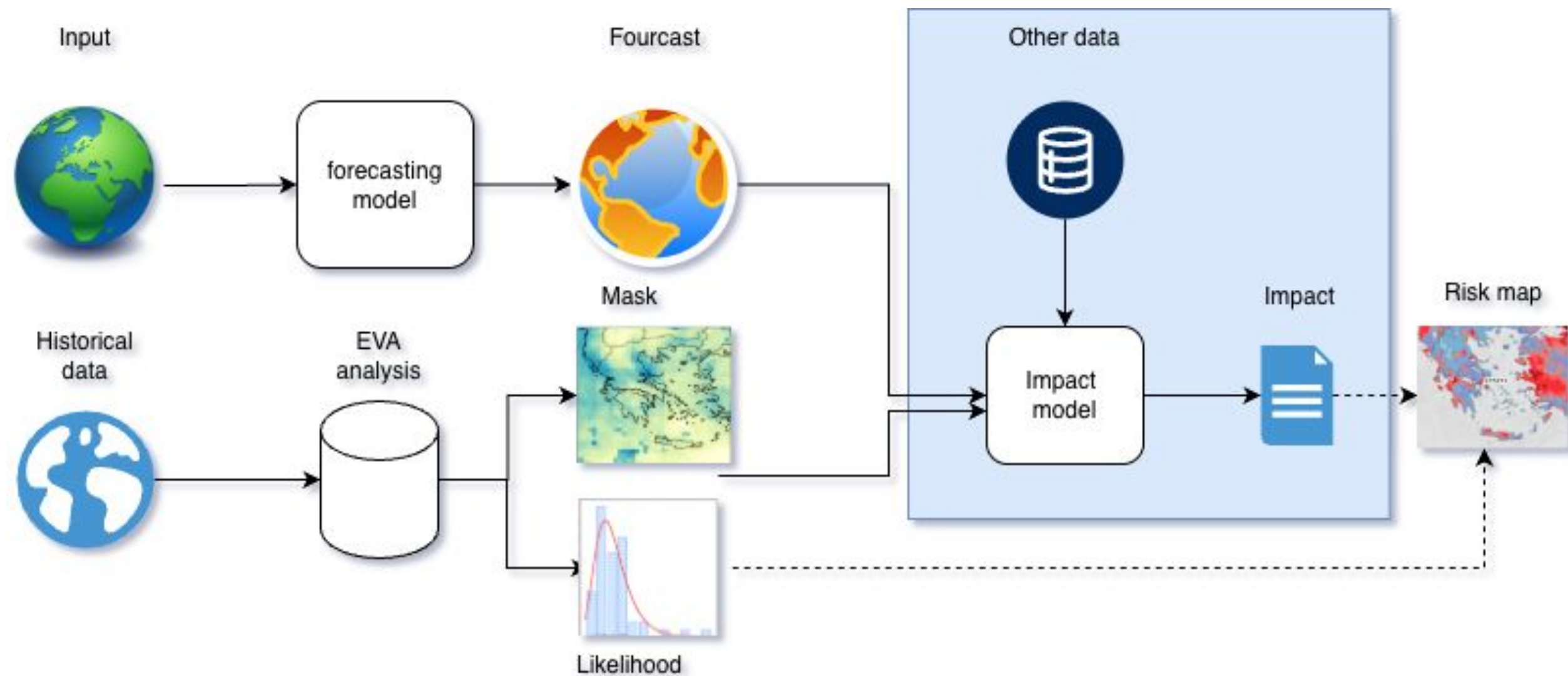
Which model is preferable?

Why AI for Impact

- Impacts not directly observable
- No simple cause–effect link
- Many factors shape impact
- Highly heterogeneous data sources
- **AI learns hidden relationships**

PART 2

Regional Flood Impact Prediction



Scope of This Work

INPUT

Weather data

ERA5 precipitation

Climate variables

Köppen–Geiger classes

Socio-economics

Population, density, GDP

MODEL

ML algorithms

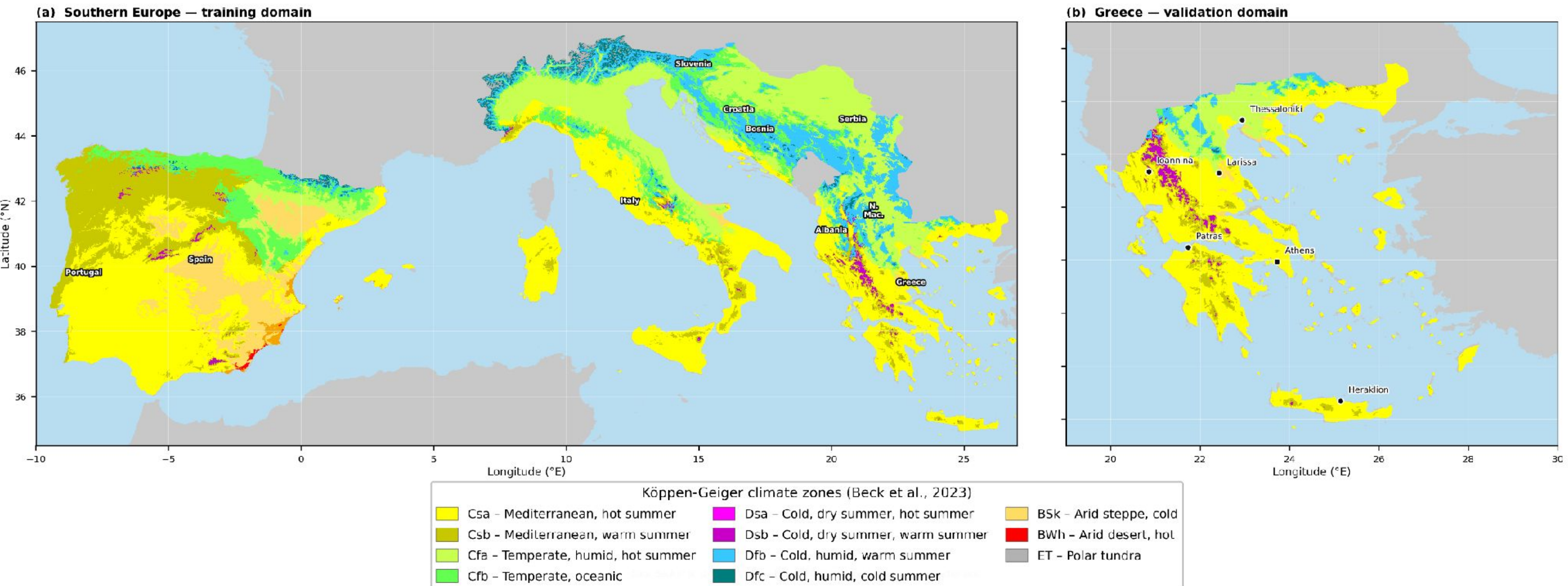
Tree ensembles, boosting,
linear

OUTPUT

Impact factor

$$IF = \frac{\text{affected population}}{\Sigma \text{ NUTS2 population}}$$

Study Domain

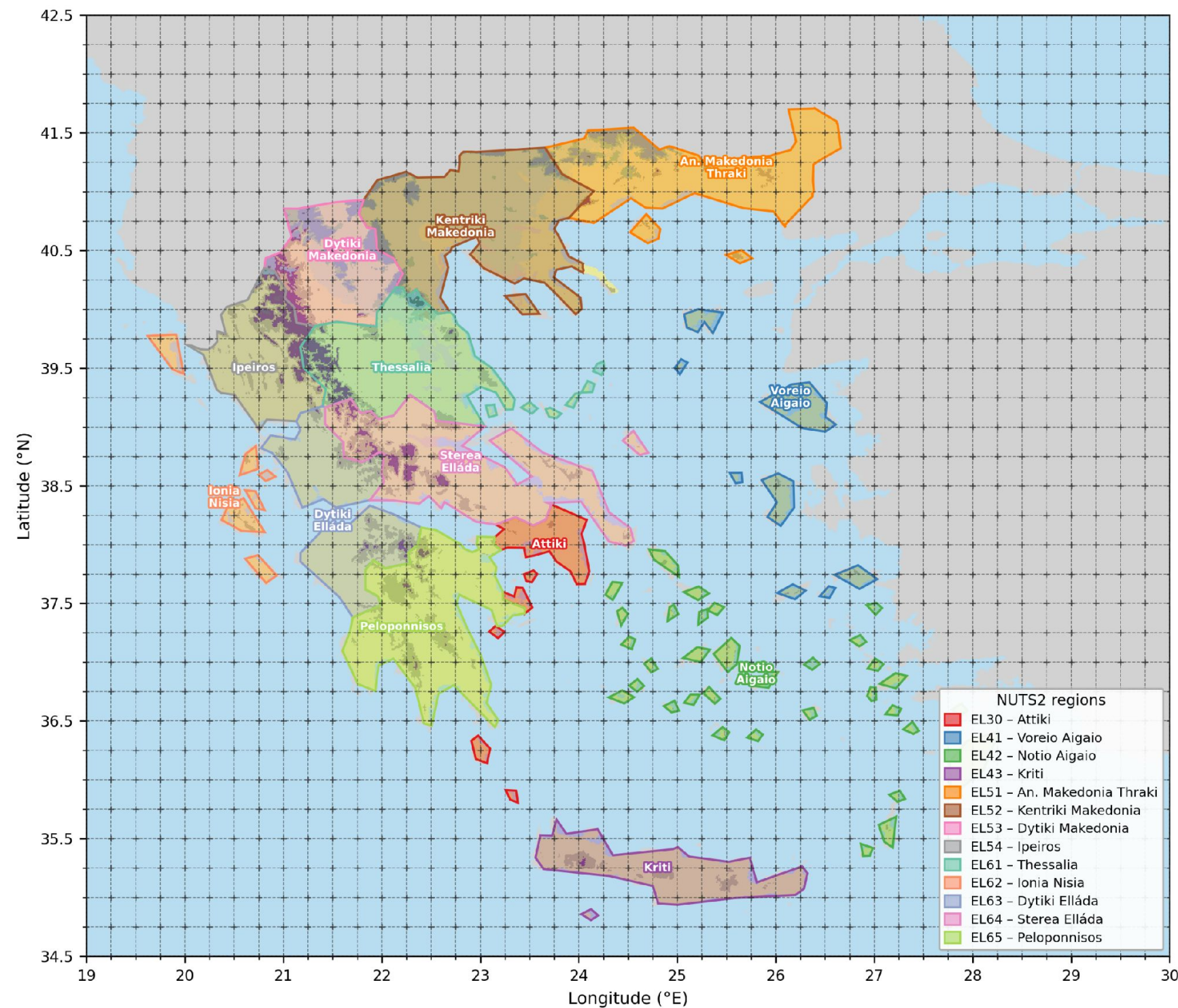


10 countries · 2000–2026

Mediterranean + continental climates

Greece withheld from training

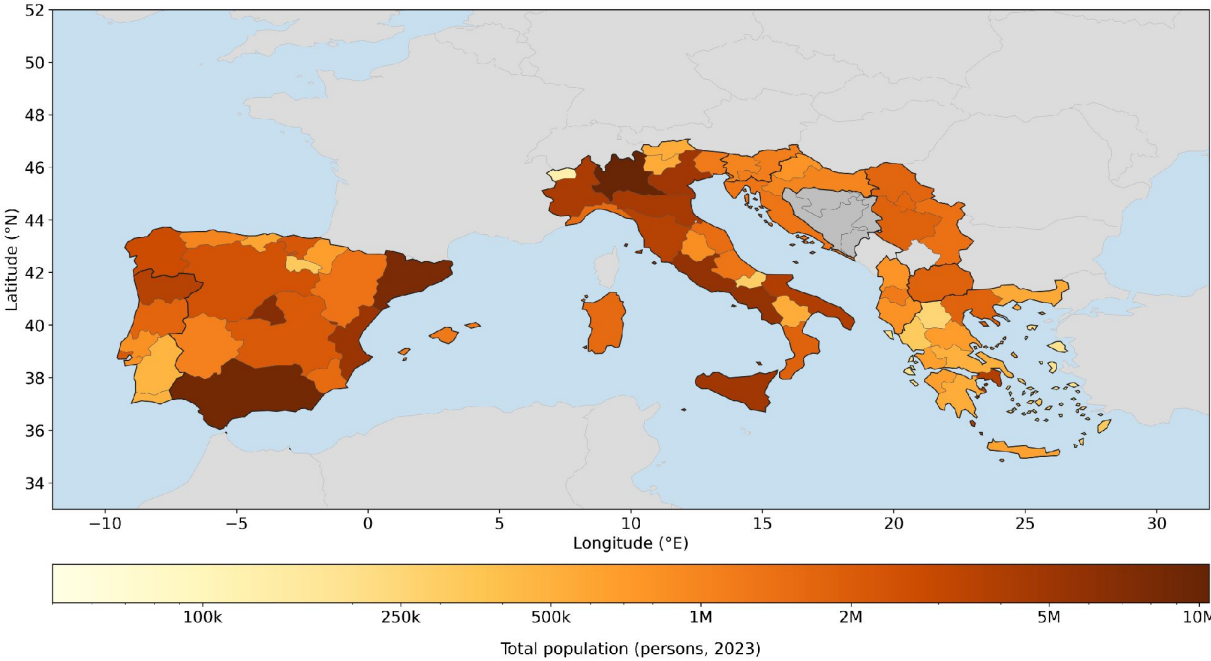
NUTS2 Regions and ERA5 Grid



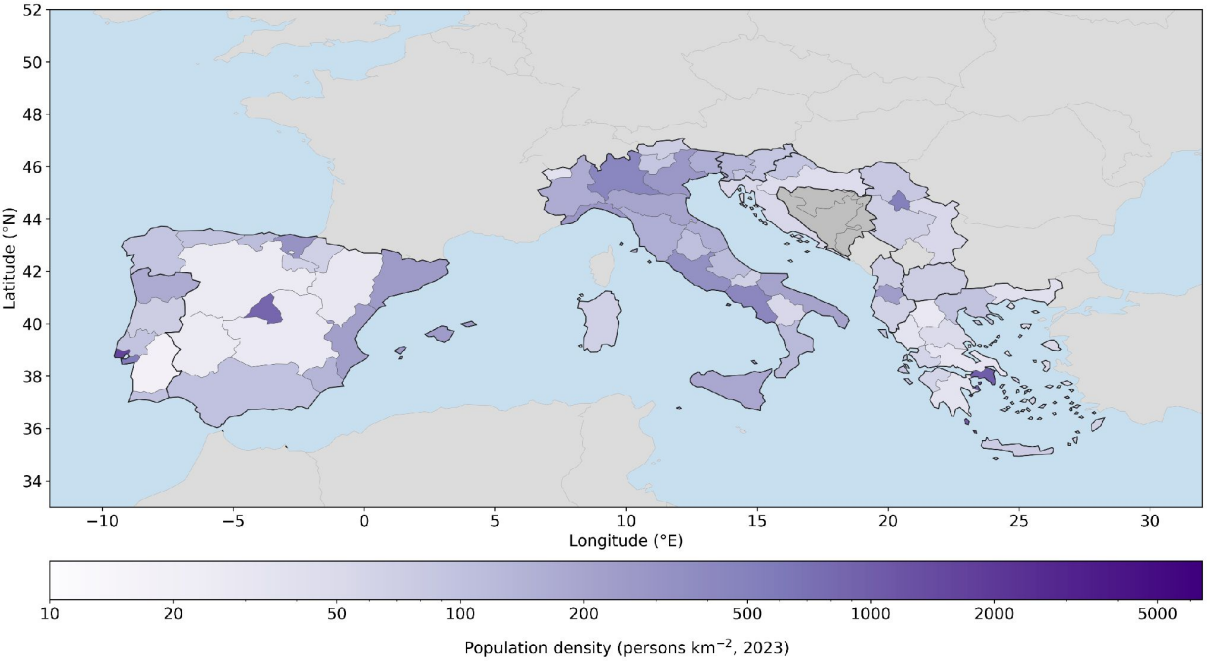
- 13 Greek NUTS2 regions
- ERA5 grid: $0.25^\circ \times 0.25^\circ$
- Spatial mean inside polygon
- Islands: few grid cells
- Uneven spatial sampling

Socio-Economic Context

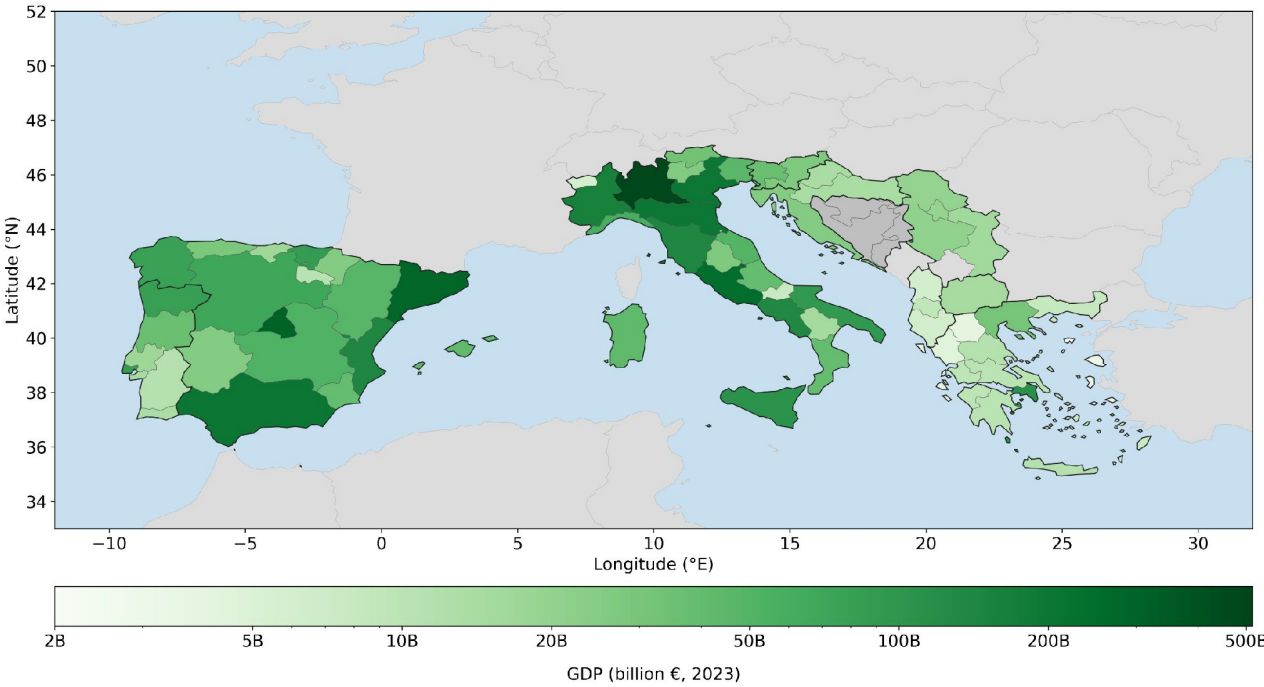
Total population



Population density

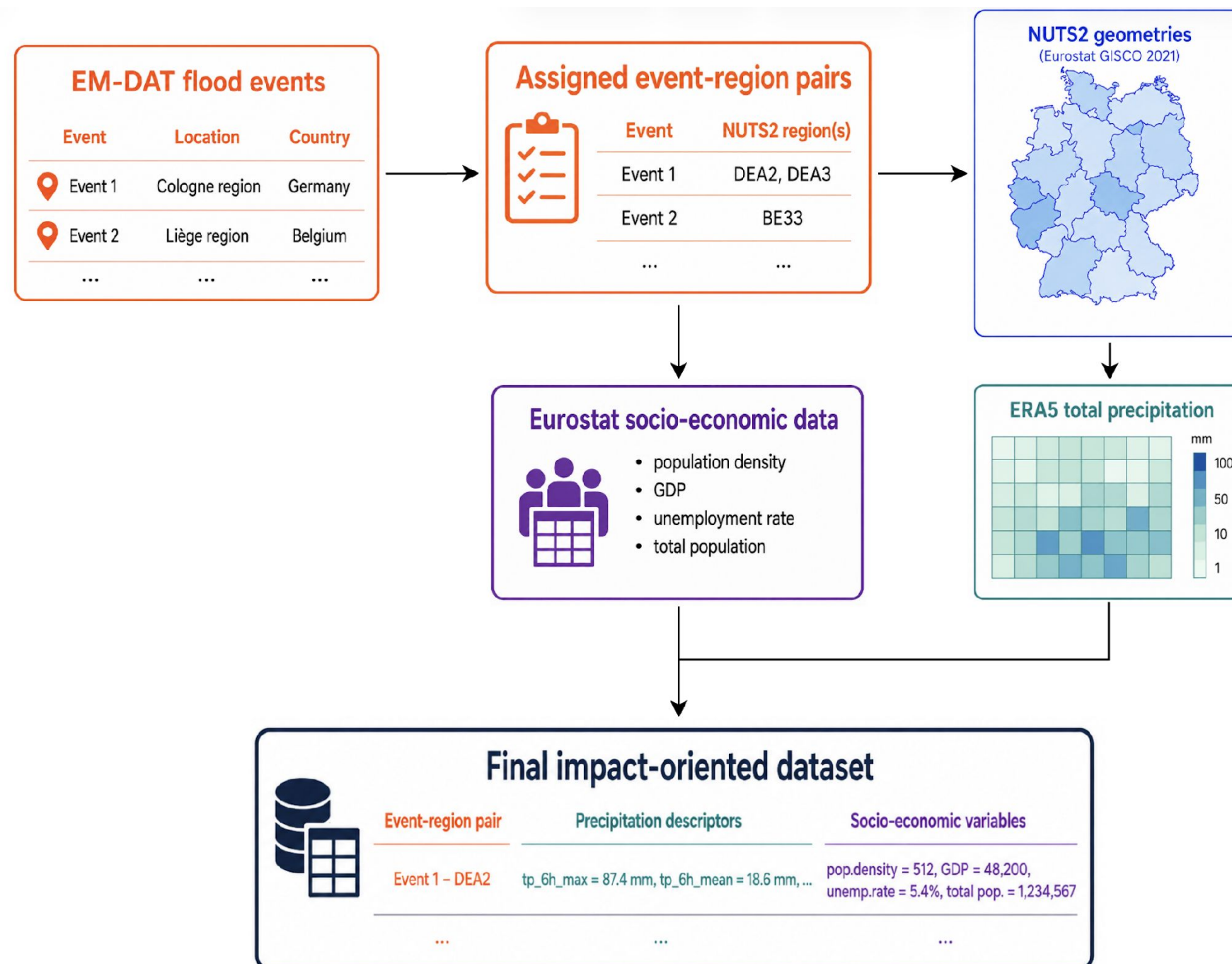


GDP (M EUR)

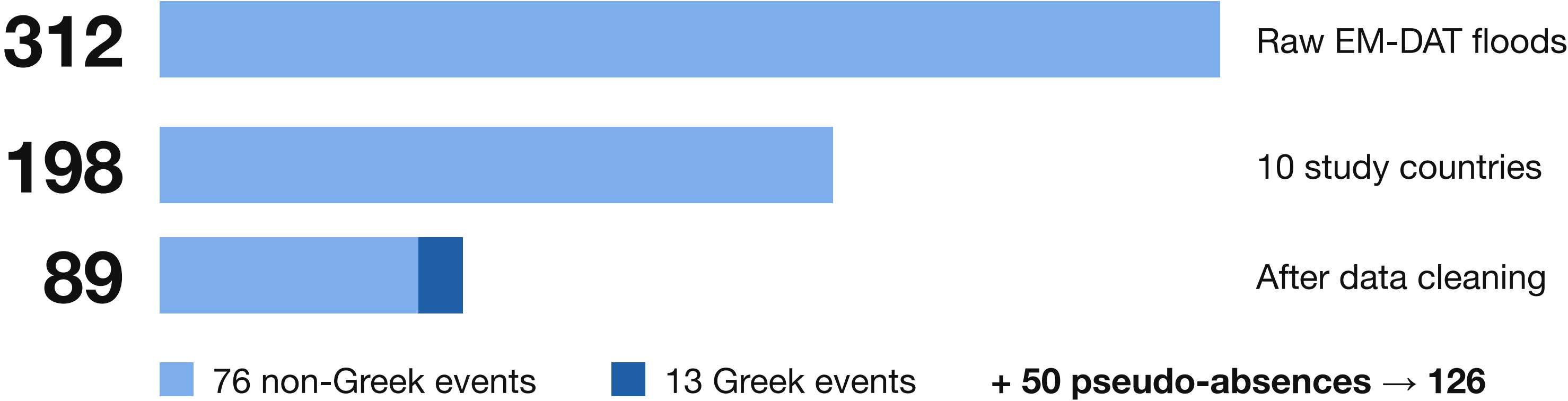


- Eurostat, NUTS2 level
- Fixed 2023 snapshot
- Strong west–east GDP gradient

Dataset Integration



- EM-DAT flood events
- ERA5 hourly tp
- Köppen–Geiger, 1 km
- Eurostat NUTS2 statistics
- **Key: event–region pairs**



Two-Stage Pipeline



- Stages trained independently
- Chained at inference
- Stage 2: real events only

Greece withheld from training

Stage 1 Results



Impact vs no impact · Greek holdout n = 21

Model	CV AUC	Val Acc	Val F1	Val AUC
RF	0.875	0.762	0.800	0.880
Extra Trees	0.880	0.810	0.833	0.933
HGB	0.753	0.762	0.783	0.841
XGBoost	0.838	0.762	0.800	0.865
LightGBM	0.774	0.762	0.783	0.856
CatBoost	0.866	0.810	0.833	0.923
Logistic Reg.	0.794	0.810	0.818	0.856

0.933

Extra Trees · Val AUC

- Val AUC: 0.84–0.93
- Linear baseline: AUC 0.856
- Strong, largely separable signal

Low vs high impact · Greek holdout n = 13

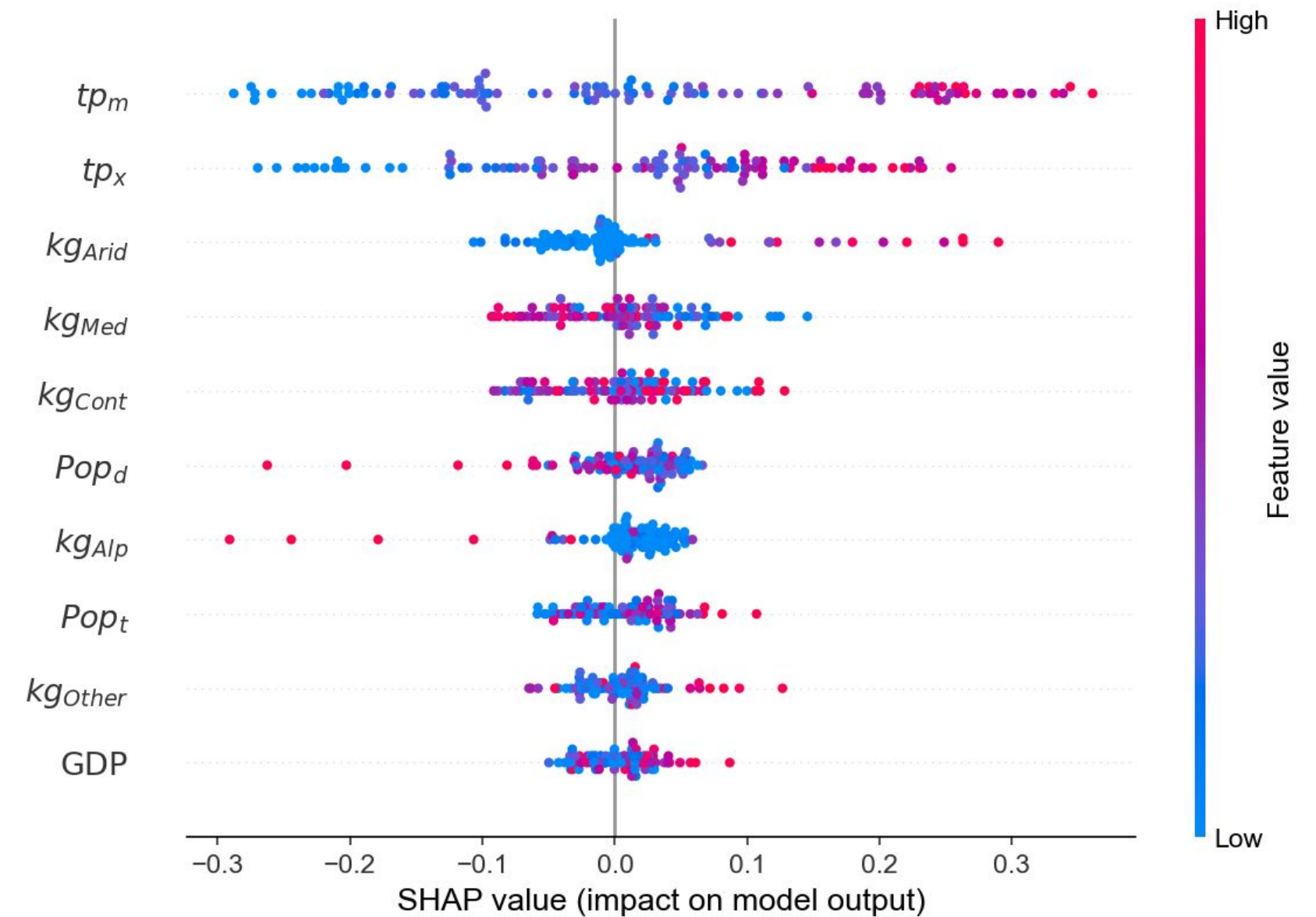
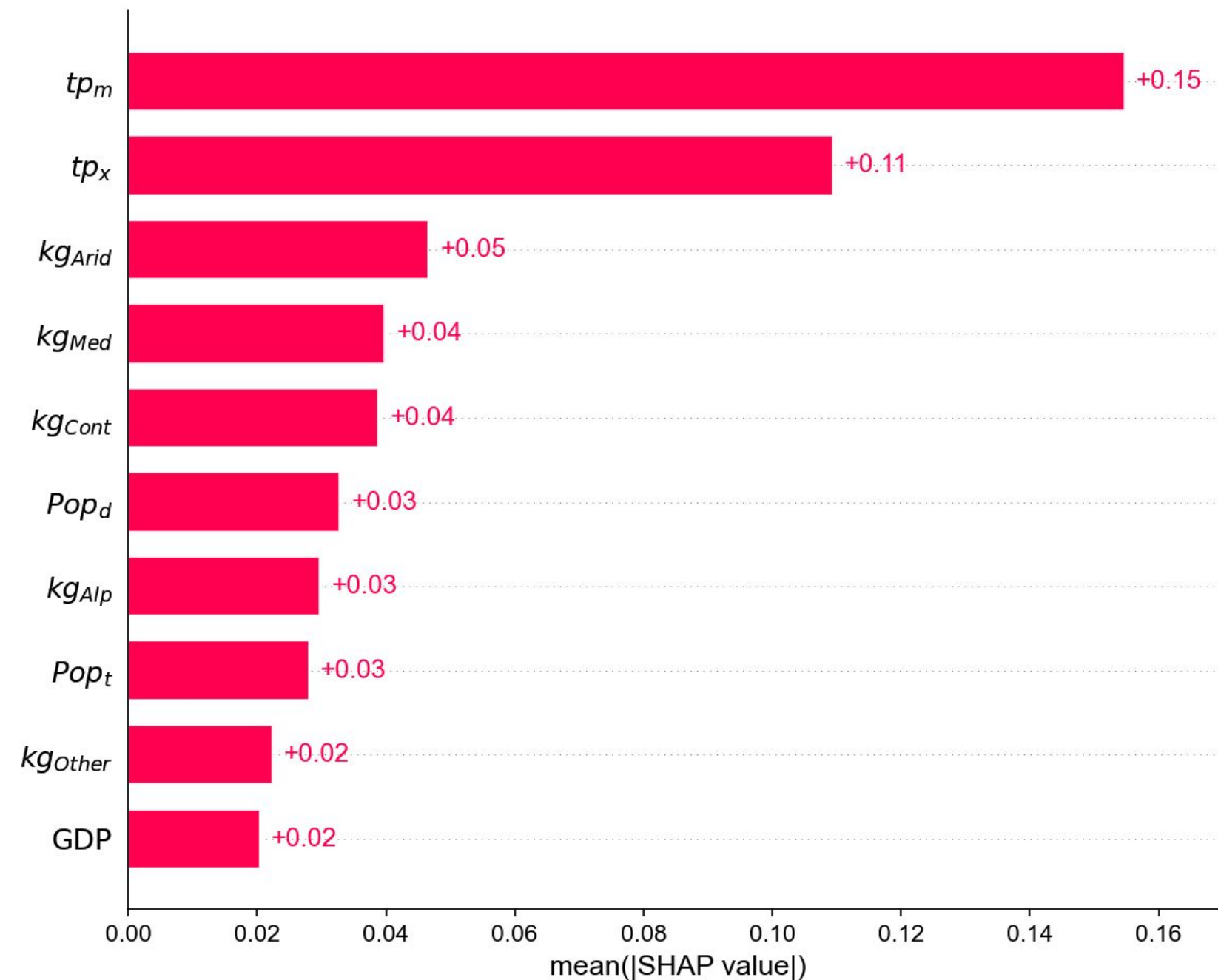
Model	CV AUC	Val Acc	Val F1	Val AUC
RF	0.623	0.923	0.960	0.833
Extra Trees	0.639	0.923	0.960	0.833
HGB	0.684	0.615	0.762	0.667
XGBoost	0.672	0.692	0.818	0.667
LightGBM	0.693	0.692	0.818	0.333
CatBoost	0.602	0.846	0.917	0.417
Logistic Reg.	0.633	0.769	0.870	0.000

1 : 12

Low : high in holdout

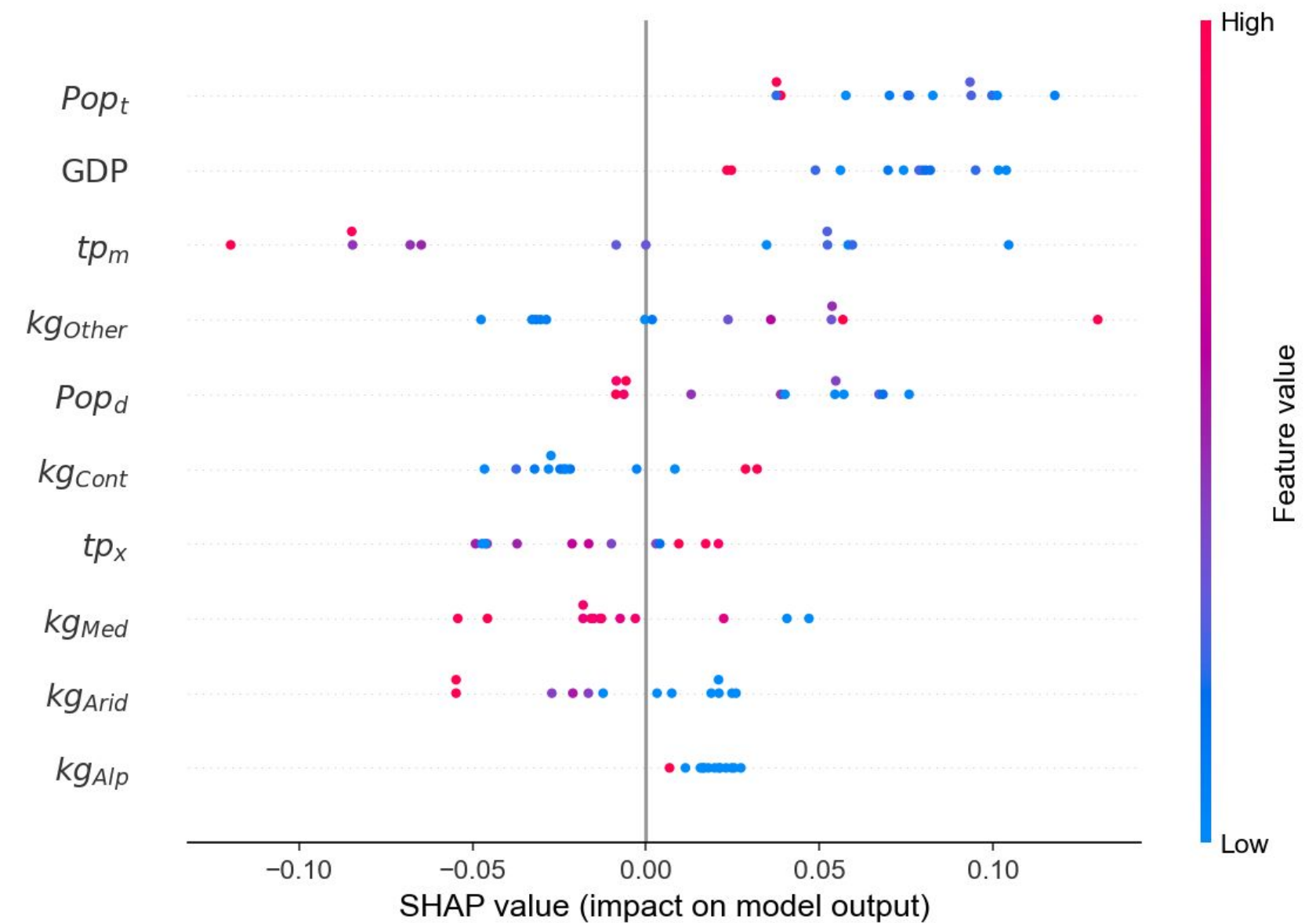
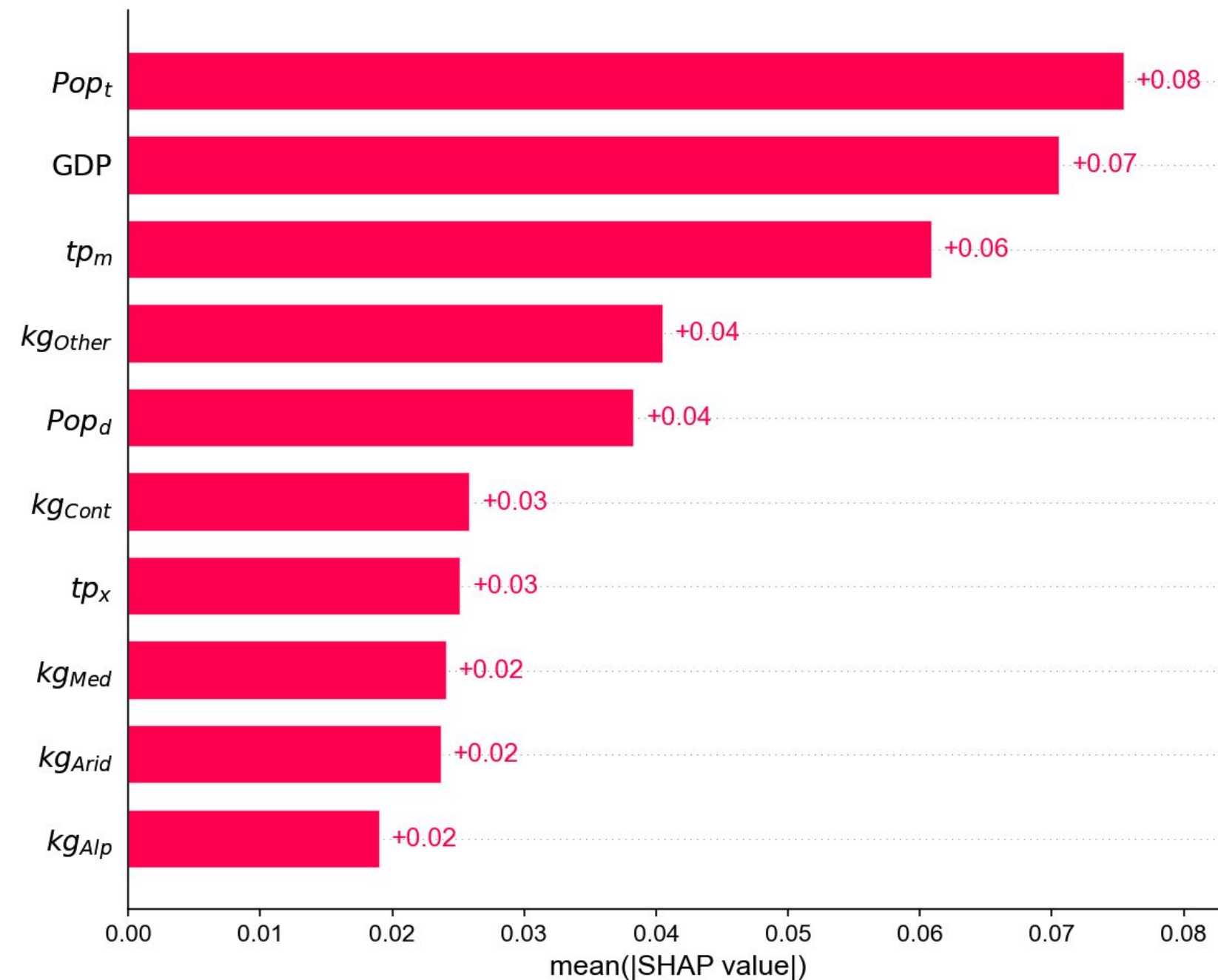
- AUC is primary metric
- All-high baseline: 0.923 Acc
- CV AUC only 0.60–0.69
- LightGBM: 0.693 → 0.333

Stage 1 Feature Attribution



- tp_m , $tp_x \approx$ half of attribution
- kg_{arid} ranks third
- Socio-economic: smallest individually

Stage 2 Feature Attribution

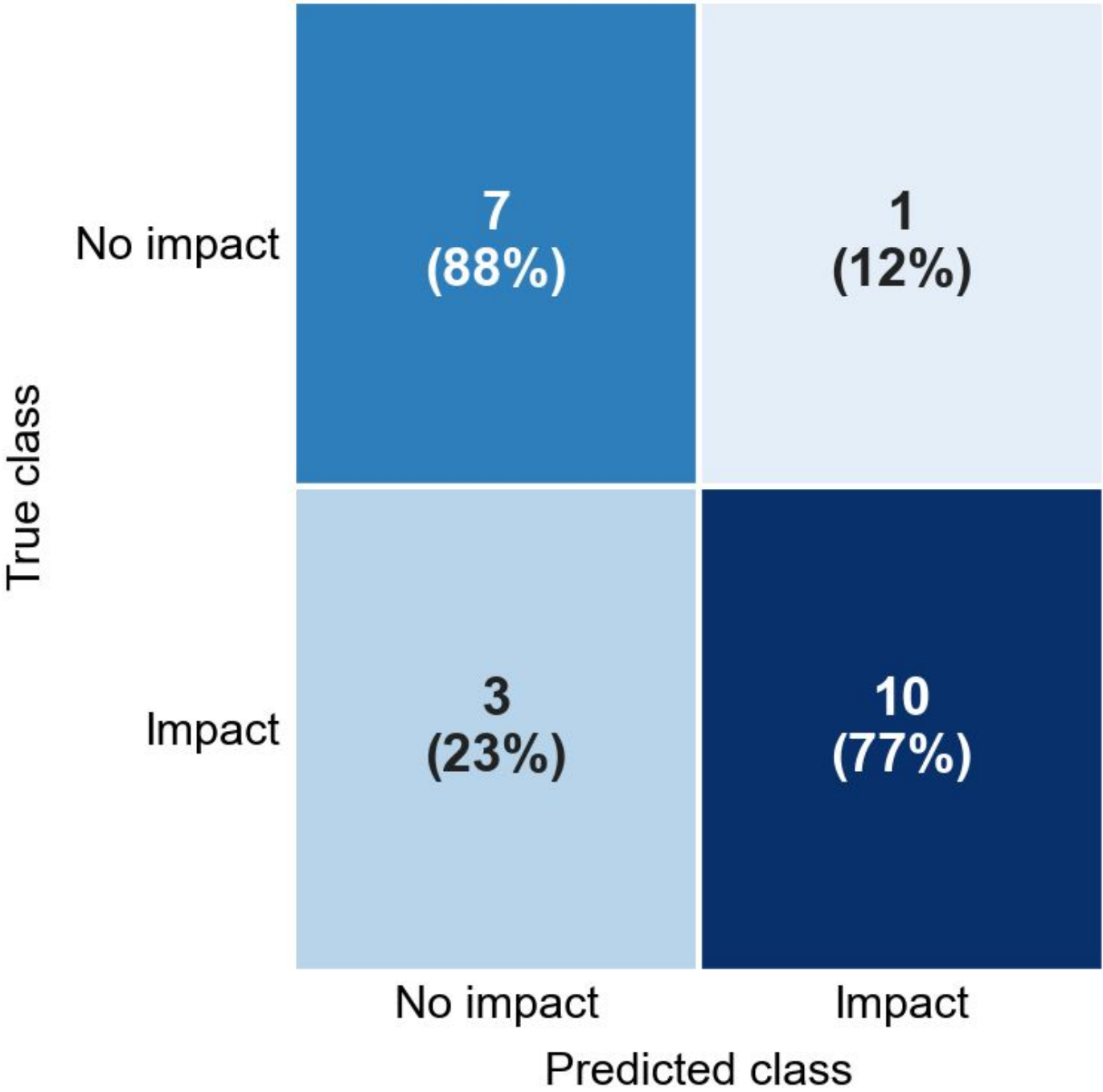


- pop_total, gdp lead
- tpm direction reversed
- Flatter importance profile

Confusion Matrices

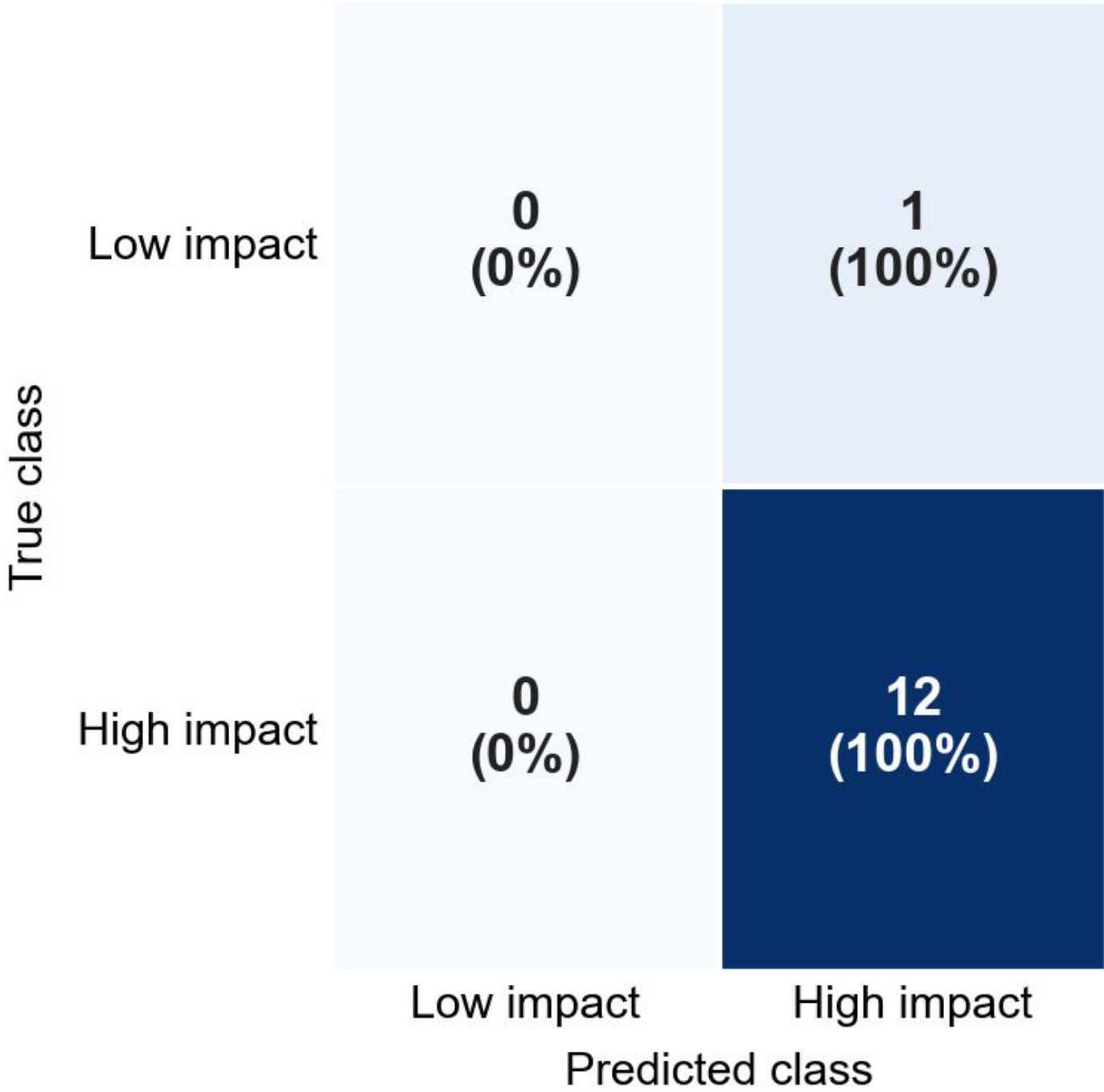
Stage 1 · n = 21

17 / 21 correct



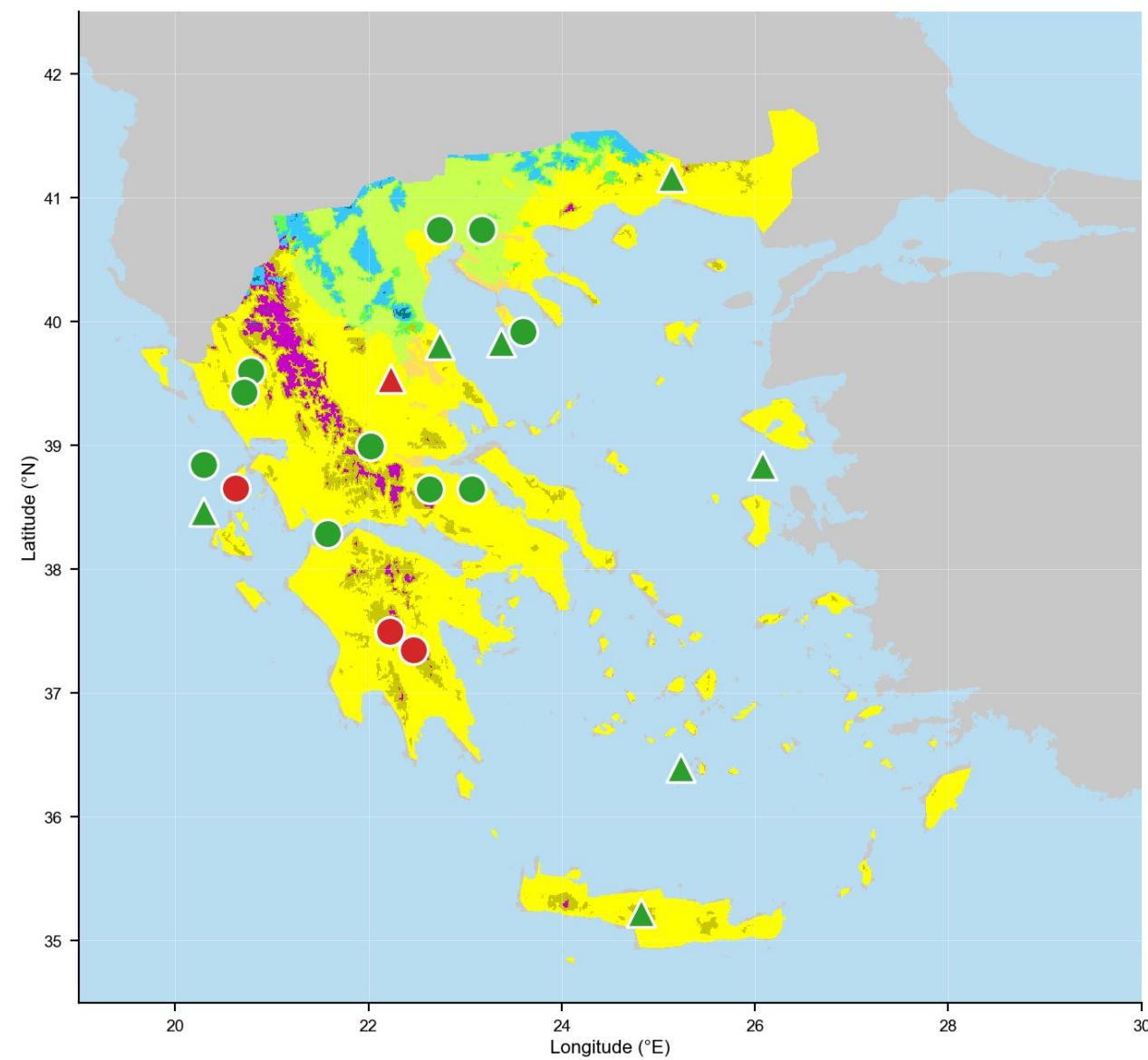
Stage 2 · n = 13

12 / 13 correct

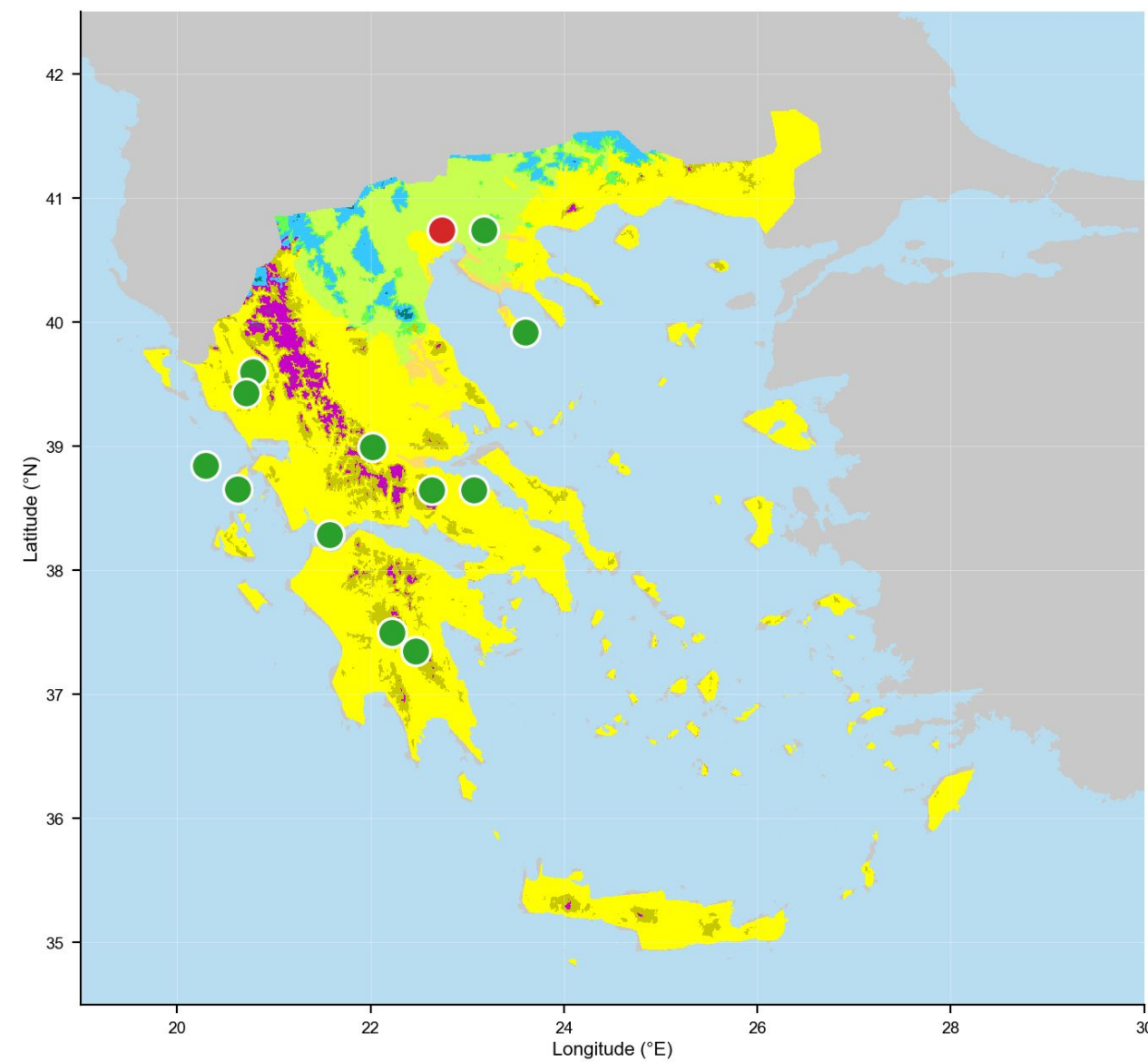


Spatial Error Analysis

Stage 1



Stage 2



- Correct
 - Misclassified
 - Real event
 - Pseudo-absence
- S1 misses: EL62, EL65
 - S2 miss: EL52, 2016

ITERATION 1

Pseudo-absences

Dry season, low rain

ML algorithms

7 classifiers

Stage 1 · Val AUC

0.933

Stage 2 · Val AUC

0.833



ITERATION 2

Pseudo-absences

New, closer to problem

ML algorithms

+ TabICL

Tabular foundation model

Results

?

DONE

First iteration

Two-stage pipeline, Greek holdout

IN PROGRESS

Writing the paper

FUTURE WORK

- Expand the impact dataset
- Improve severity classification
- Add hydrological variables
- Improve spatial matching

Conclusions

METHODOLOGY STATUS

DONE

EVA masks

Extreme thresholds for Greece

DONE

Forecasting

Extreme precipitation benchmark

DONE

Impact model

Stage 1: reliable detection
Stage 2: needs more data

FUTURE

Risk maps

Likelihood + impact

PART 1

FUTURE WORK

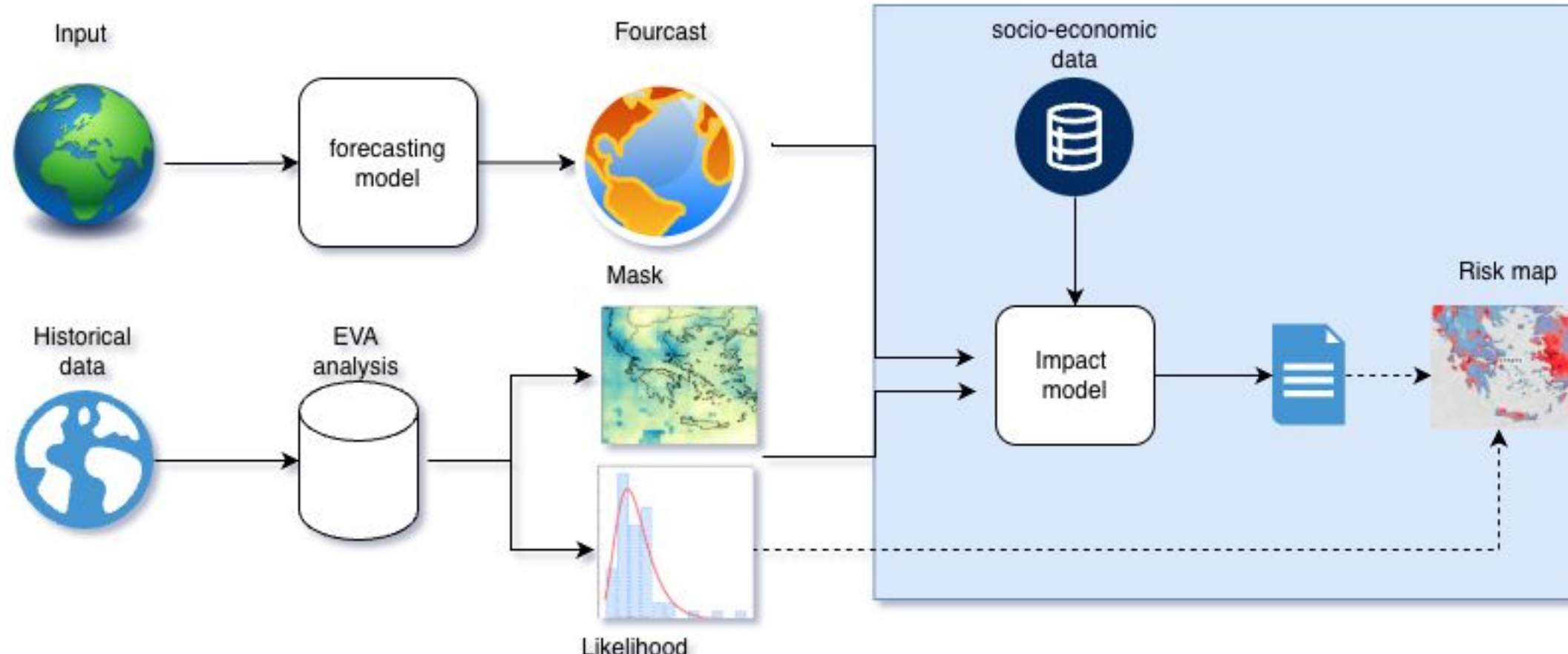
1 Expand the dataset

2 Economic losses: agriculture

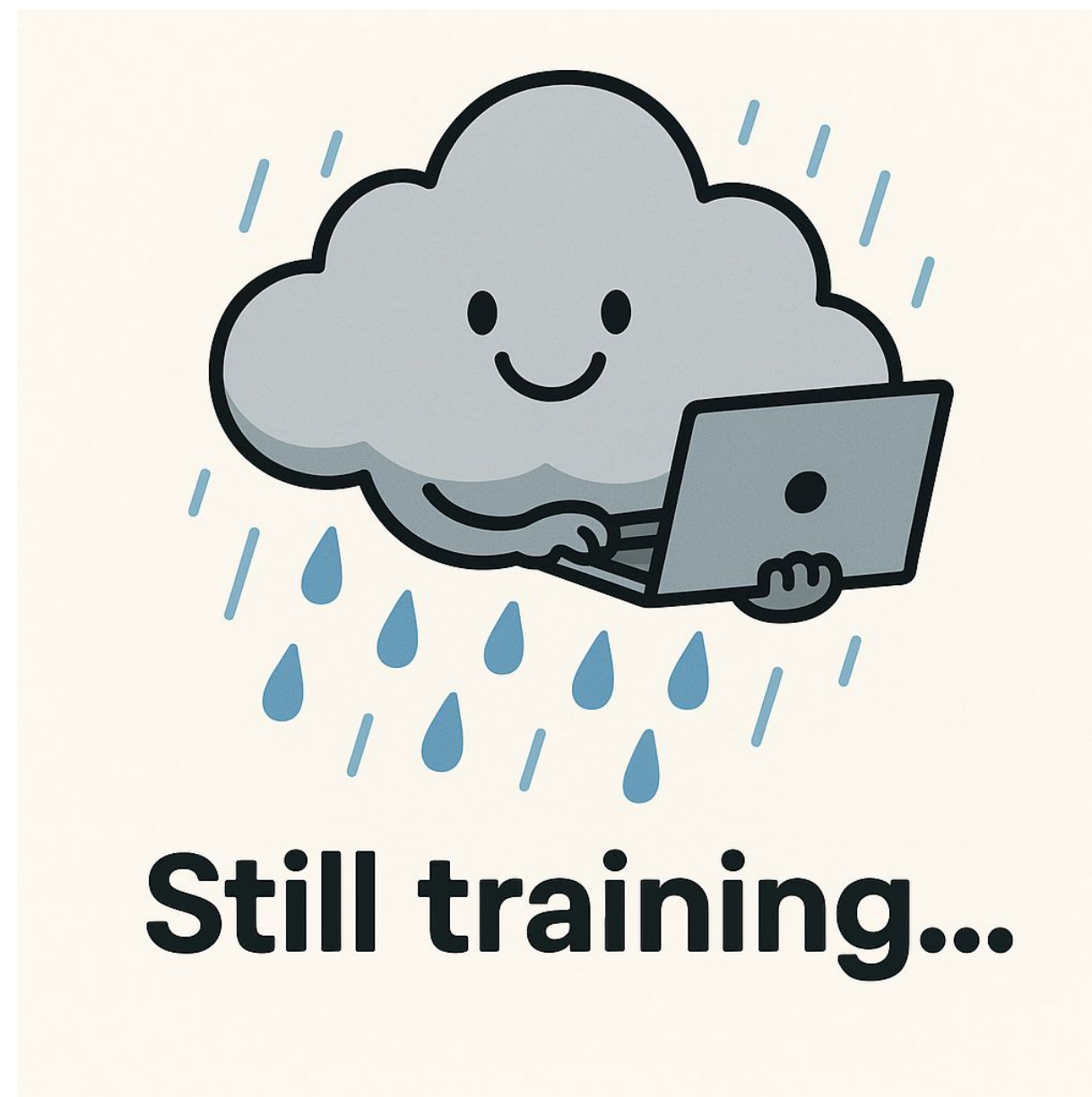
3 Test full methodology

PART 2

MISSING



Thank you very much for your attention!



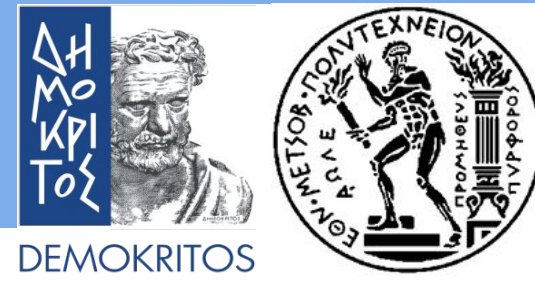
vasileios.vatellis@iit.demokritos.gr

APPENDIX

Backup Slides

Study	Hazard	Impact proxy	Unit	Method
Ritter et al., 2021	Floods, Europe	Exposed population	Pan-EU grid	Rule-based (EFAS)
Iglesias et al., 2024	Severe weather	Emergency calls	Municipality	Random Forest
Morales et al., 2026	Floods, Catalonia	Emergency calls	Municipality	XGBoost
Kooshki Forooshani et al., 2024	Tropical cyclones	House damage	0.1° grid	XGBoost, two-stage
Zheng et al., 2025	Tropical cyclones	Economic loss	County	CatBoost + SHAP
This work	Floods, S. Europe	Affected pop. ratio	NUTS2	7 classifiers, two-stage

References



- Pathak et al., 2022, Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators [arXiv: 2202.11214](#)
- Bi et al., 2023 , Accurate medium-range global weather forecasting with 3d neural networks. [Doi: 10.1038/s41586-023-06185-3](#)
- Lam et al., 2023, Learning skillful medium-range global weather forecasting. [Doi: 10.1126/science.adi2336](#)
- Price et al., 2025, Probabilistic weather forecasting with machine learning. [Doi: 10.1038/s41586-024-08252-9](#)
- Chen et al., 2023, FuXi: a cascade machine learning forecasting system for 15-day global weather forecast. Doi: 10.1038/s41612-023-00512-1
- Bodnar et al., 2025, A foundation model for the Earth system. Doi: 10.1038/s41586-025-09005-y

Precipitation

ERA5, per event

Peak precipitation value per grid point,
over time

Mean accumulation, over NUTS2
region

Climate

Köppen–Geiger fractions

kg_med

kg_cont

kg_arid

kg_alp

kg_other

Socio-economic

Eurostat, NUTS2

pop_total

Total population

pop_den

Population density

gdp

GDP, million EUR

1

More impact records

2

NUTS3 or catchment scale

3

Hydrological and terrain features

4

Island and continental coverage

5

Economic loss target

Comparison with Related Work

Kooshki Forooshani et al., 2024

0.648

F1 · cyclone damage

- Precipitation dominates detection

This work · Stage 1

0.833

F1 · Greek holdout

- Low data limits all

Morales et al., 2026

28.8%

CSI · dense municipalities

- Not directly comparable

- Small, imbalanced sample
- Pseudo-absences are assumed
- ERA5 smooths local extremes
- Static 2023 socio-economics
- Reported dates may shift

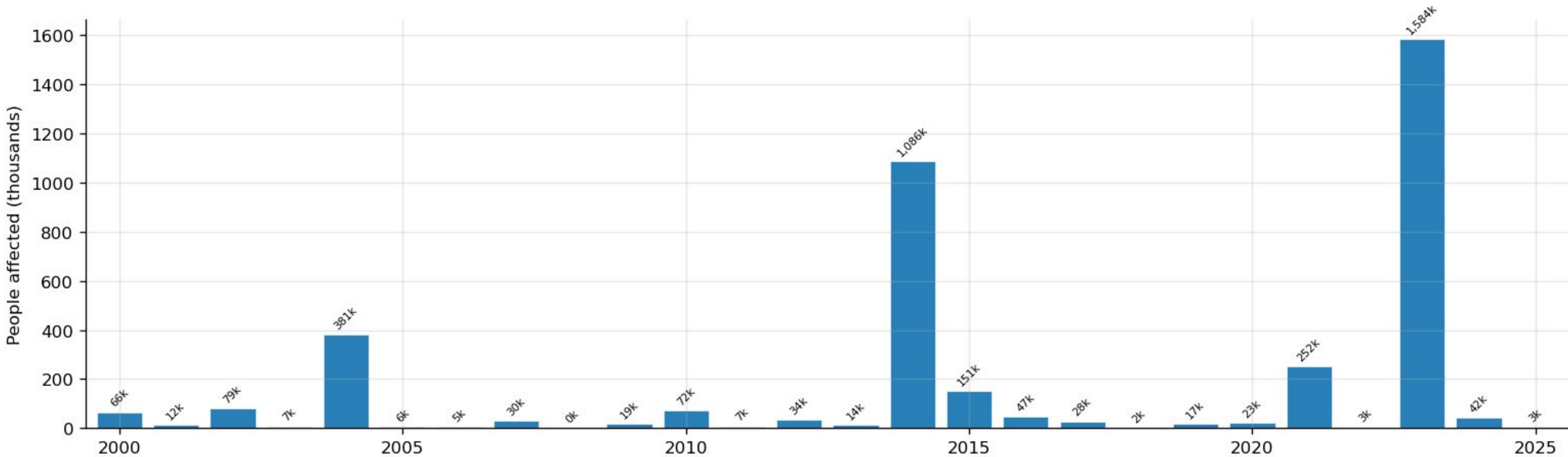
126

samples in total

76 real · 50 pseudo-absence

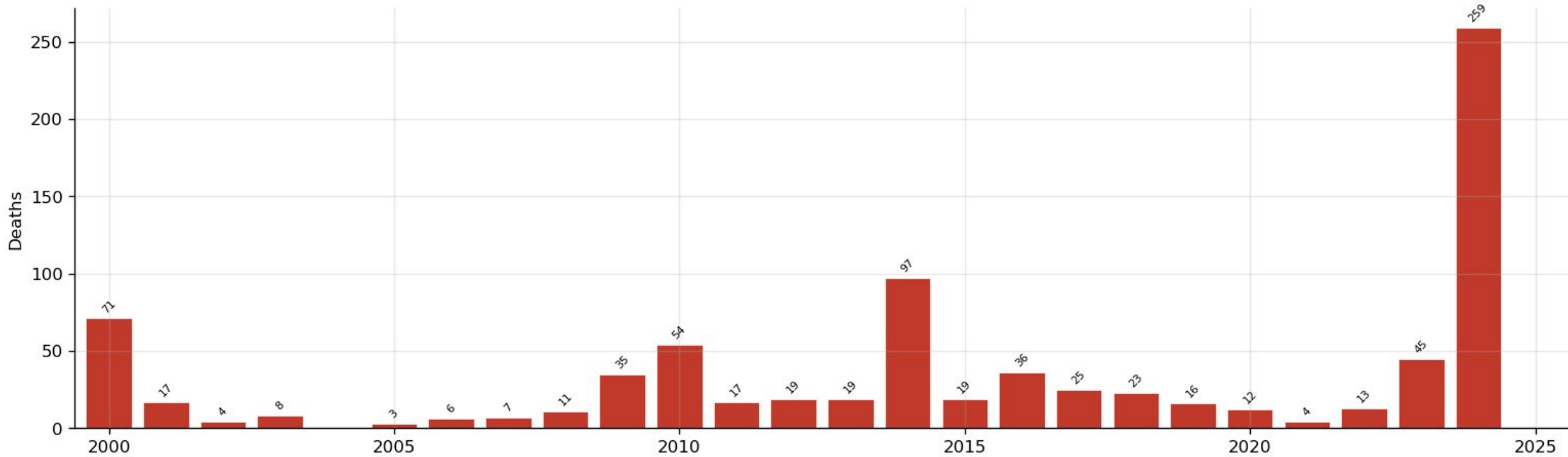
Flood Impact in Southern Europe

Affected people
per year



Flood deaths
per year

EM-DAT · 10 countries ·
2000–2026



STAGE 1 · EL65

Peloponnese

Orographic dilution

- Taygetos, Parnon trigger rainfall
- Area mean hides peaks

STAGE 1 · EL62

Ionian Islands

Out of distribution

- No island training regions
- $kg_other \approx 23\%$

STAGE 2 · EL52

Central Macedonia

Continental gap

- 2016 low-impact event
- Predicted as high

Pseudo-Absence Samples

TYPE A · DRY SEASON

Random July–August dates

TYPE B · LOW RAINFALL

$\text{tpm} < 0.208 \text{ mm h}^{-1}$ ($\approx 5 \text{ mm day}^{-1}$)

- ± 30 days from known events
- Country share mirrors positives

Median ERA5 descriptors (mm)

	Real events	Pseudo-abs.
tpx	32.45	6.51
tpm	3.19	0.15

Pseudo-absences are not all dry days

Training and Holdout Sets

STAGE 1 · TRAIN

105

63 real events
42 pseudo-absences

STAGE 2 · TRAIN

63

32 low impact
31 high impact

GREEK HOLDOUT

21

13 real: 1 low, 12 high
8 pseudo-absences

13 non-Greek events excluded from Stage 2: missing affected population

- Default hyperparameters, no tuning
- 5-fold stratified CV
- Balanced class weights
- Retrain on all non-Greek
- Single test on Greece

BAGGING

Random Forest

Extra Trees

BOOSTING

HistGradBoost

XGBoost

LightGBM

CatBoost

LINEAR BASELINE

Logistic Regression

STAGE 1 · EXTRA TREES

Greek holdout Val AUC

All 10 features **0.933**



Top-2 only (tpm, tpx) **0.885**



Drop tpm **0.875**



Precipitation carries most signal

STAGE 2 · EXTRA TREES

Greek holdout Val AUC

All 10 features **0.833**



Top-2 only (pop_total, gdp) **0.958**



Drop pop_total **0.833**



Top-2 CV AUC only 0.573

Stage 1 Full Results



CV: mean $\pm$ sd, 5 folds, 105 samples · Val: Greek holdout, 21 samples

Model	CV Acc	CV F1	CV AUC	Val Acc	Val F1	Val AUC
RF	0.771 $\pm$ 0.085	0.808 $\pm$ 0.082	0.875 $\pm$ 0.074	0.762	0.800	0.880
ET	0.810 $\pm$ 0.058	0.845 $\pm$ 0.049	0.880 $\pm$ 0.087	0.810	0.833	0.933
HGB	0.657 $\pm$ 0.062	0.698 $\pm$ 0.083	0.753 $\pm$ 0.149	0.762	0.783	0.841
XGBoost	0.752 $\pm$ 0.071	0.787 $\pm$ 0.073	0.838 $\pm$ 0.098	0.762	0.800	0.865
LightGBM	0.686 $\pm$ 0.066	0.725 $\pm$ 0.072	0.774 $\pm$ 0.094	0.762	0.783	0.856
CatBoost	0.771 $\pm$ 0.060	0.803 $\pm$ 0.063	0.866 $\pm$ 0.081	0.810	0.833	0.923
LR	0.724 $\pm$ 0.040	0.764 $\pm$ 0.055	0.794 $\pm$ 0.088	0.810	0.818	0.856

Stage 2 Full Results



CV: mean $\pm$ sd, 5 folds, 63 samples · Val: Greek holdout, 13 samples (1 low, 12 high)

Model	CV Acc	CV F1	CV AUC	Val Acc	Val F1	Val AUC
RF	0.608 $\pm$ 0.144	0.597 $\pm$ 0.126	0.623 $\pm$ 0.115	0.923	0.960	0.833
ET	0.686 $\pm$ 0.115	0.652 $\pm$ 0.134	0.639 $\pm$ 0.169	0.923	0.960	0.833
HGB	0.671 $\pm$ 0.119	0.632 $\pm$ 0.123	0.684 $\pm$ 0.138	0.615	0.762	0.667
XGBoost	0.637 $\pm$ 0.126	0.606 $\pm$ 0.174	0.672 $\pm$ 0.125	0.692	0.818	0.667
LightGBM	0.624 $\pm$ 0.159	0.598 $\pm$ 0.177	0.693 $\pm$ 0.154	0.692	0.818	0.333
CatBoost	0.655 $\pm$ 0.187	0.632 $\pm$ 0.207	0.602 $\pm$ 0.151	0.846	0.917	0.417
LR	0.604 $\pm$ 0.119	0.589 $\pm$ 0.148	0.633 $\pm$ 0.131	0.769	0.870	0.000

- All R^2 negative
- Ridge best: $R^2 -0.571$
- XGBoost: $R^2 -14.37$
- Target strongly right-skewed
- Exposure, hydrology features missing

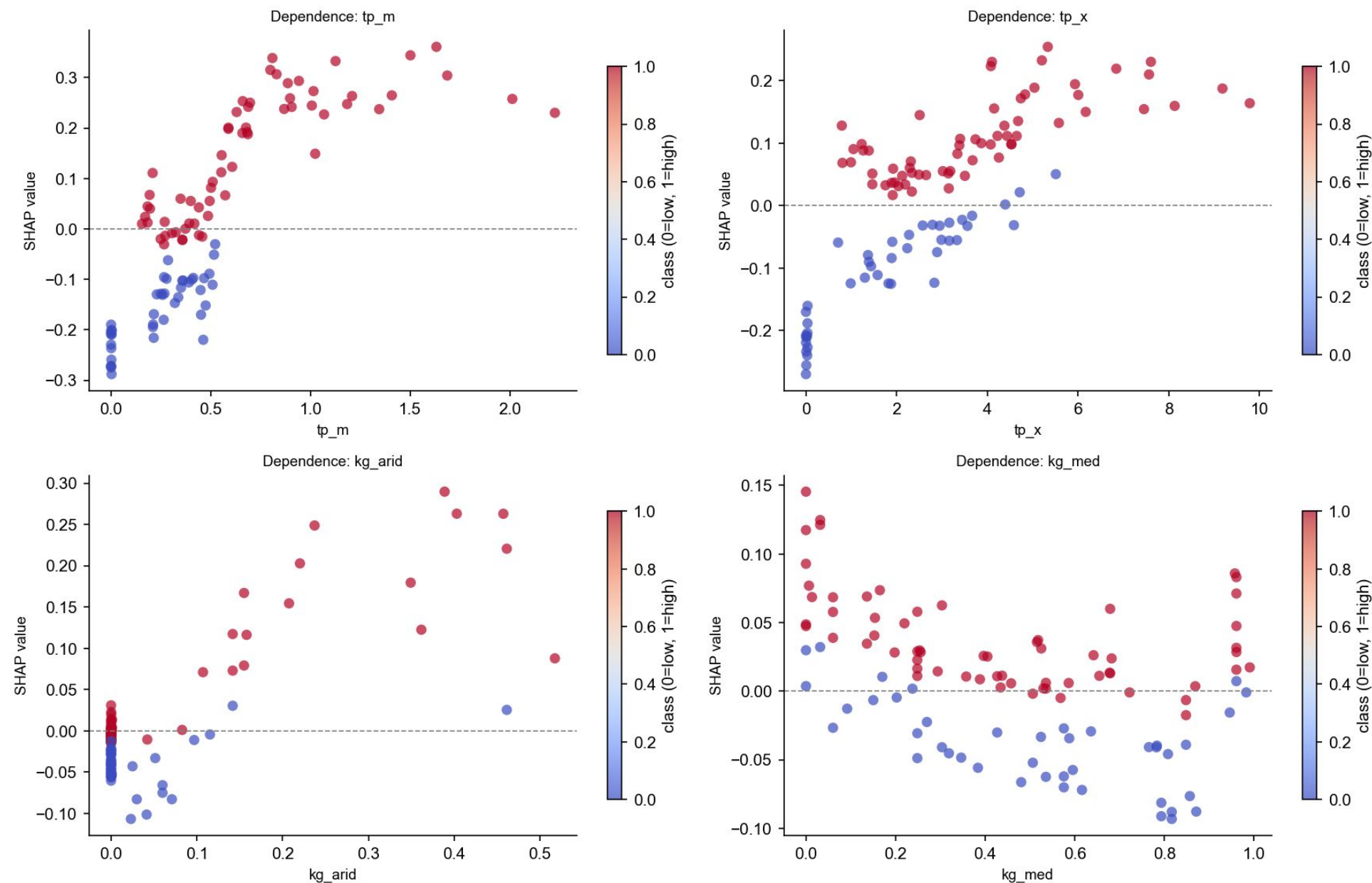
$$R^2 < 0$$

CV and Greek holdout

63 train · 13 Greek test

SHAP Dependence Plots

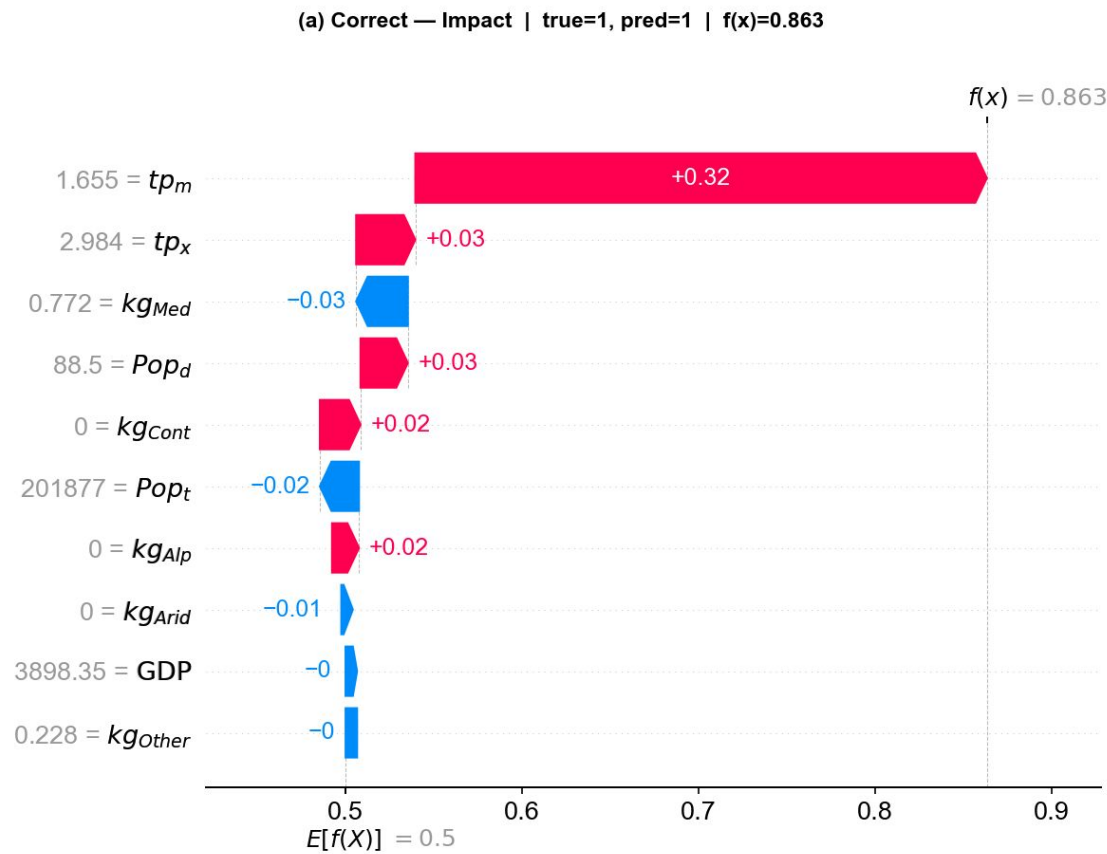
SHAP Dependence Plots — Top 4 Features (Extra Trees)



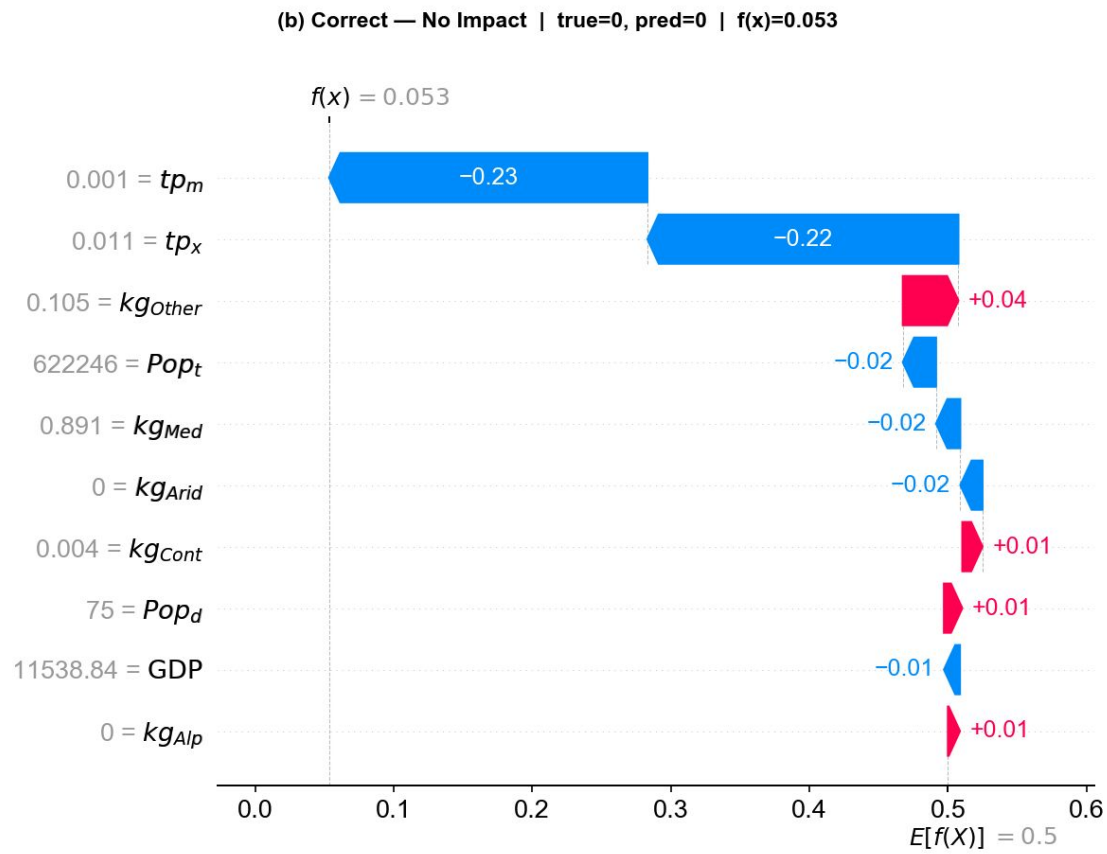
- tpm: near-monotonic increase
- Sign change $\approx 0.21 \text{ mm h}^{-1}$
- Expected by construction
- kg_arid > 0.05: positive
- kg_med: non-monotonic

SHAP Waterfall Examples

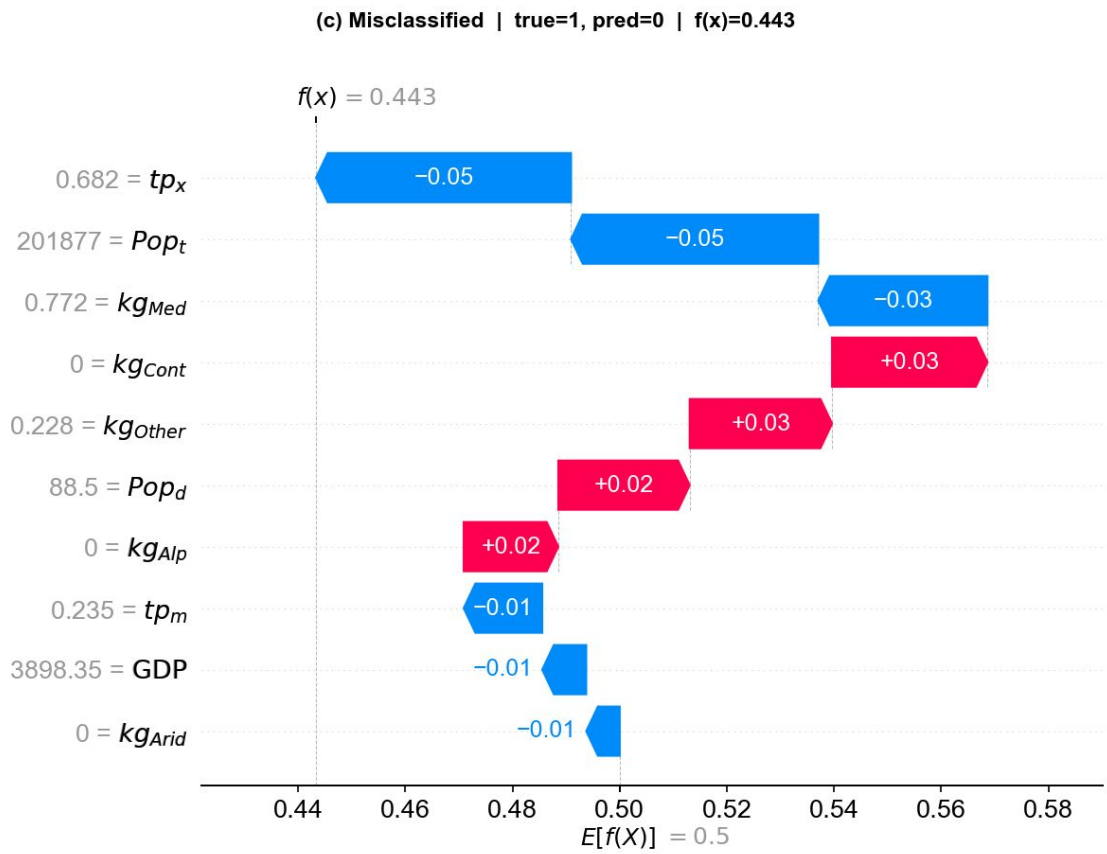
Correct · impact



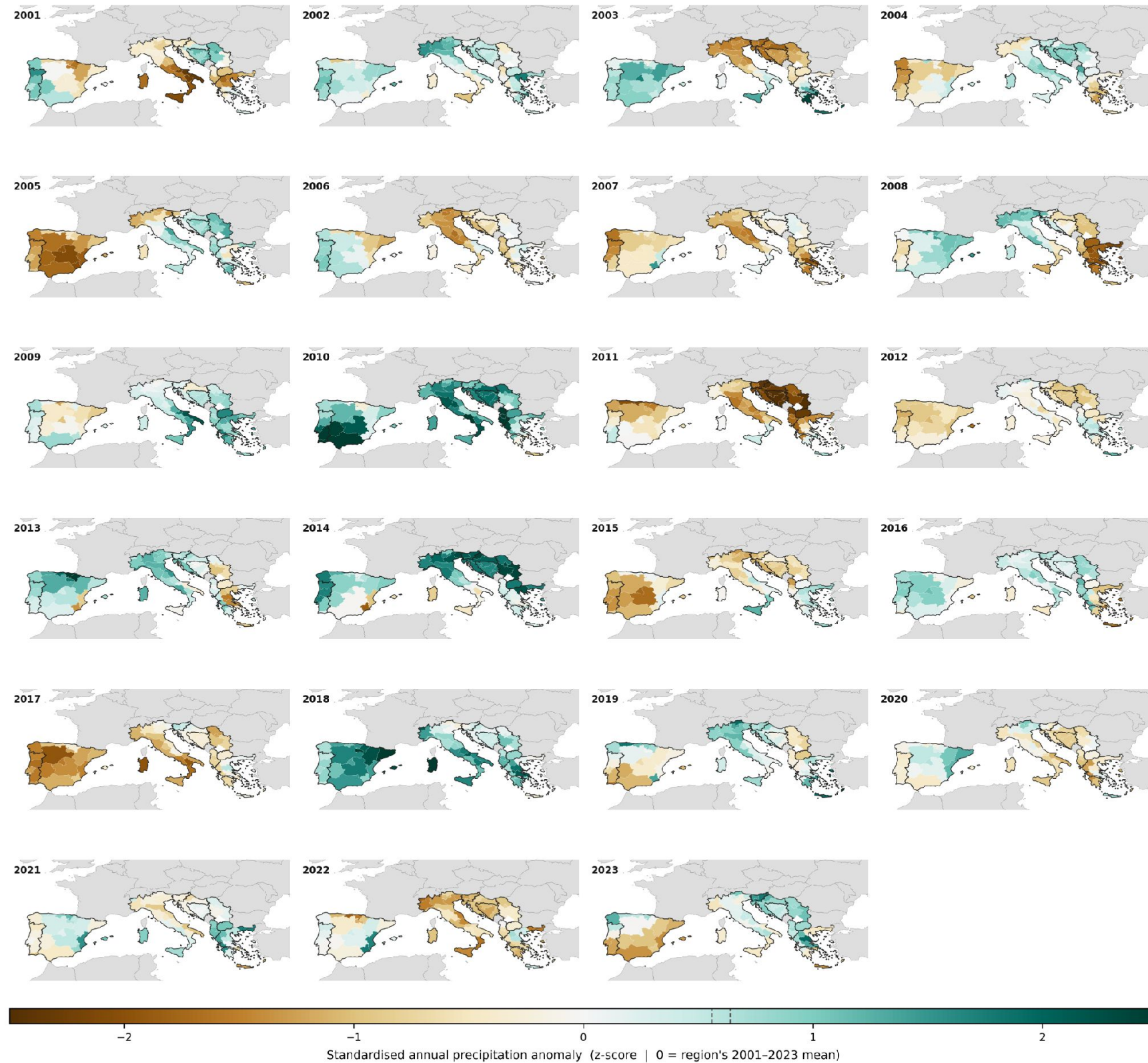
Correct · no impact



Misclassified



Precipitation Anomalies



- ERA5 z-score per NUTS2
- Wet years ↔ flood activity
- e.g. Balkans 2010, 2014