

**Data alone will not protect the environment.
Understanding them will.**



Air



Numeric data

e.g. pollutant concentrations, temperature, humidity



Photos

e.g. sky conditions, cloud cover



Satellite imagery

e.g. air quality, land use



Audio

e.g. wind, traffic, industrial noise



Text

e.g. reports, observations



Biodiversity



Photos

e.g. species identification, habitat monitoring



Audio

e.g. bird calls, acoustic monitoring



Genetic data

e.g. eDNA, barcoding



Text

e.g. species records, observations



Water



Numeric data

e.g. pH, nutrients, dissolved oxygen, temperature



Photos

e.g. water quality, underwater life



Satellite imagery

e.g. water temperature, chlorophyll



Audio

e.g. underwater noise, marine life



Text

e.g. monitoring reports, logbooks



Soil



Numeric data

e.g. soil moisture, nutrients, pH, temperature



Photos

e.g. soil structure, erosion



Satellite imagery

e.g. soil moisture, land use



Audio

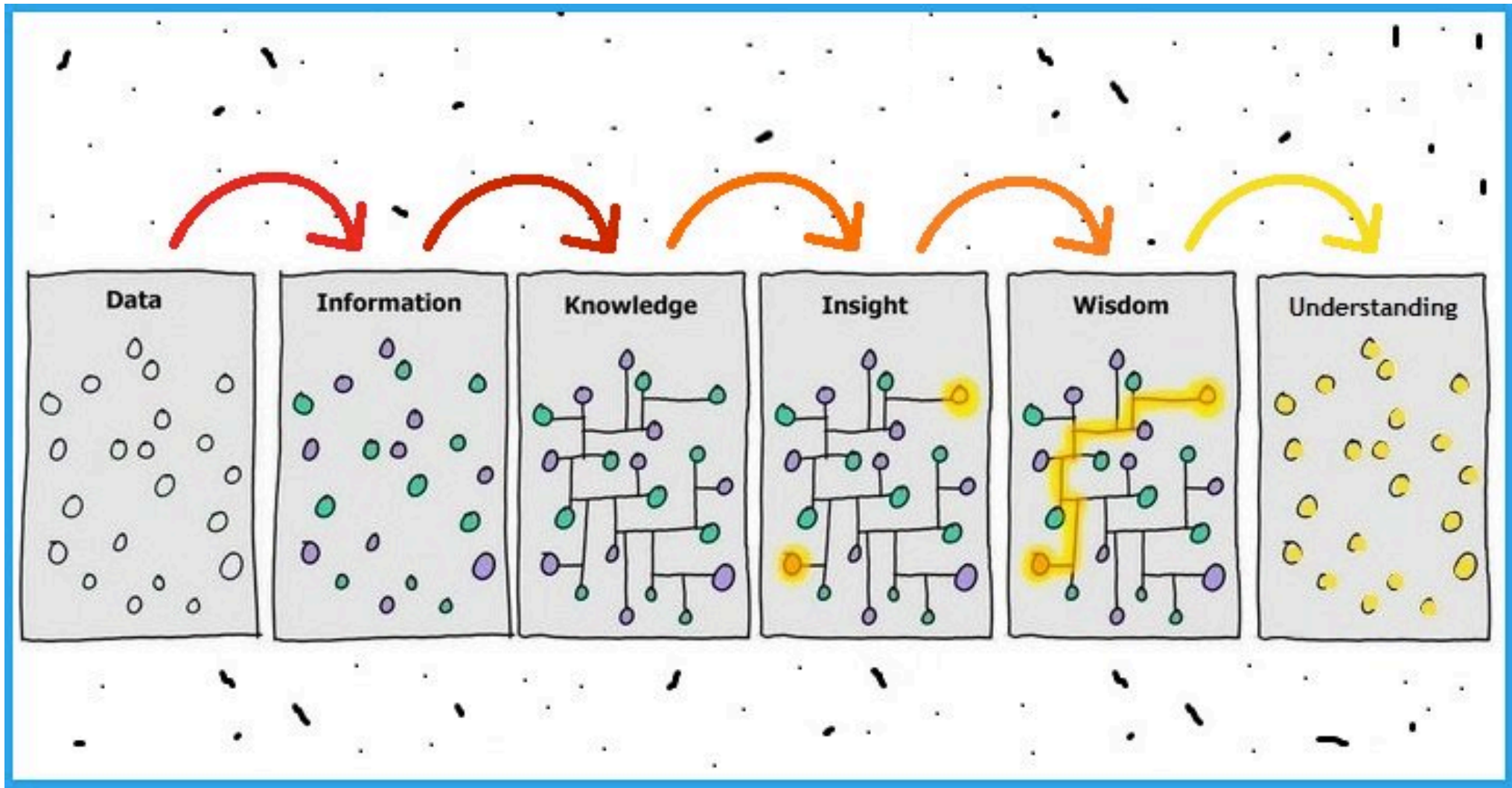
e.g. soil fauna, root activity

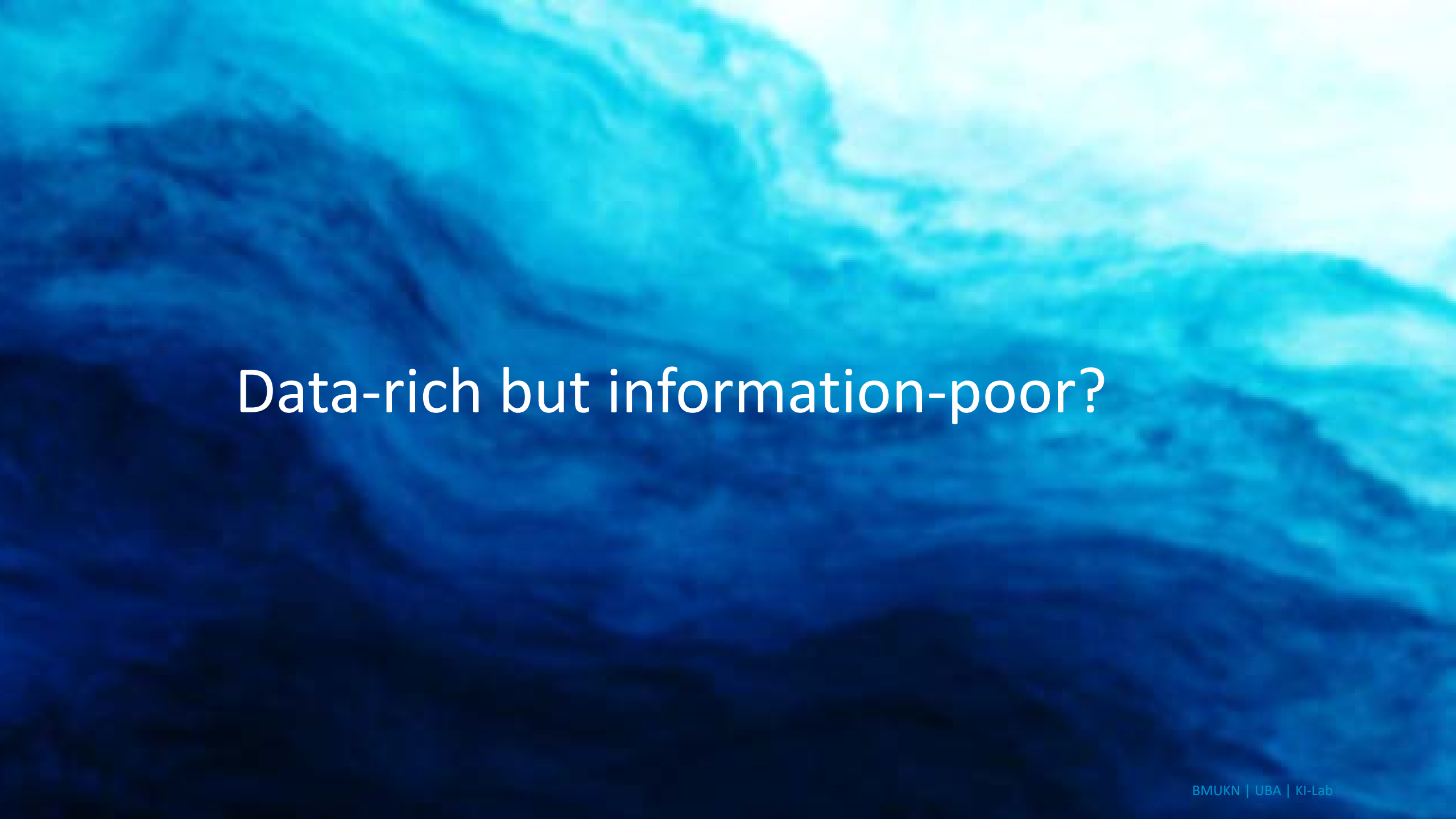


Text

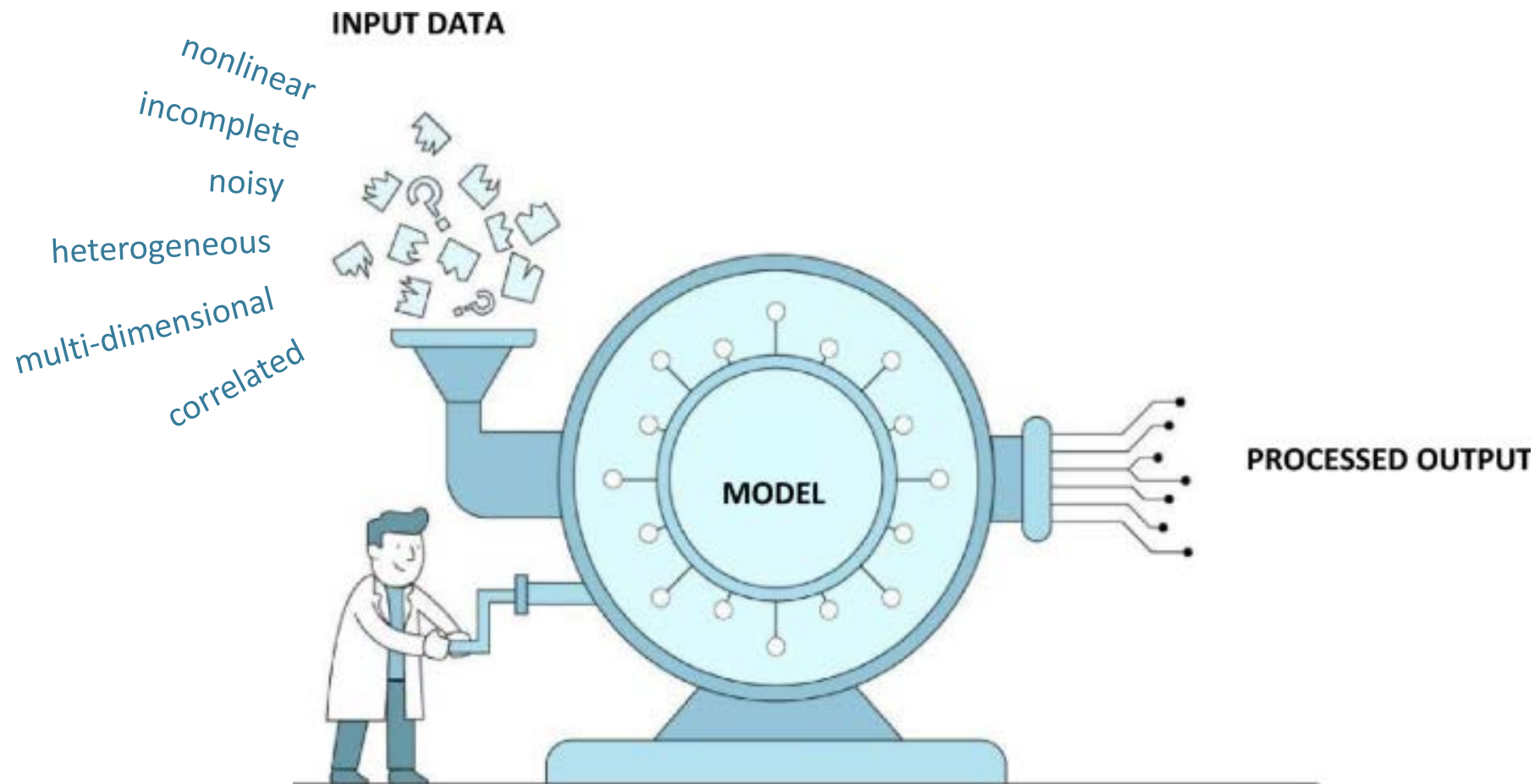
e.g. soil surveys, research papers



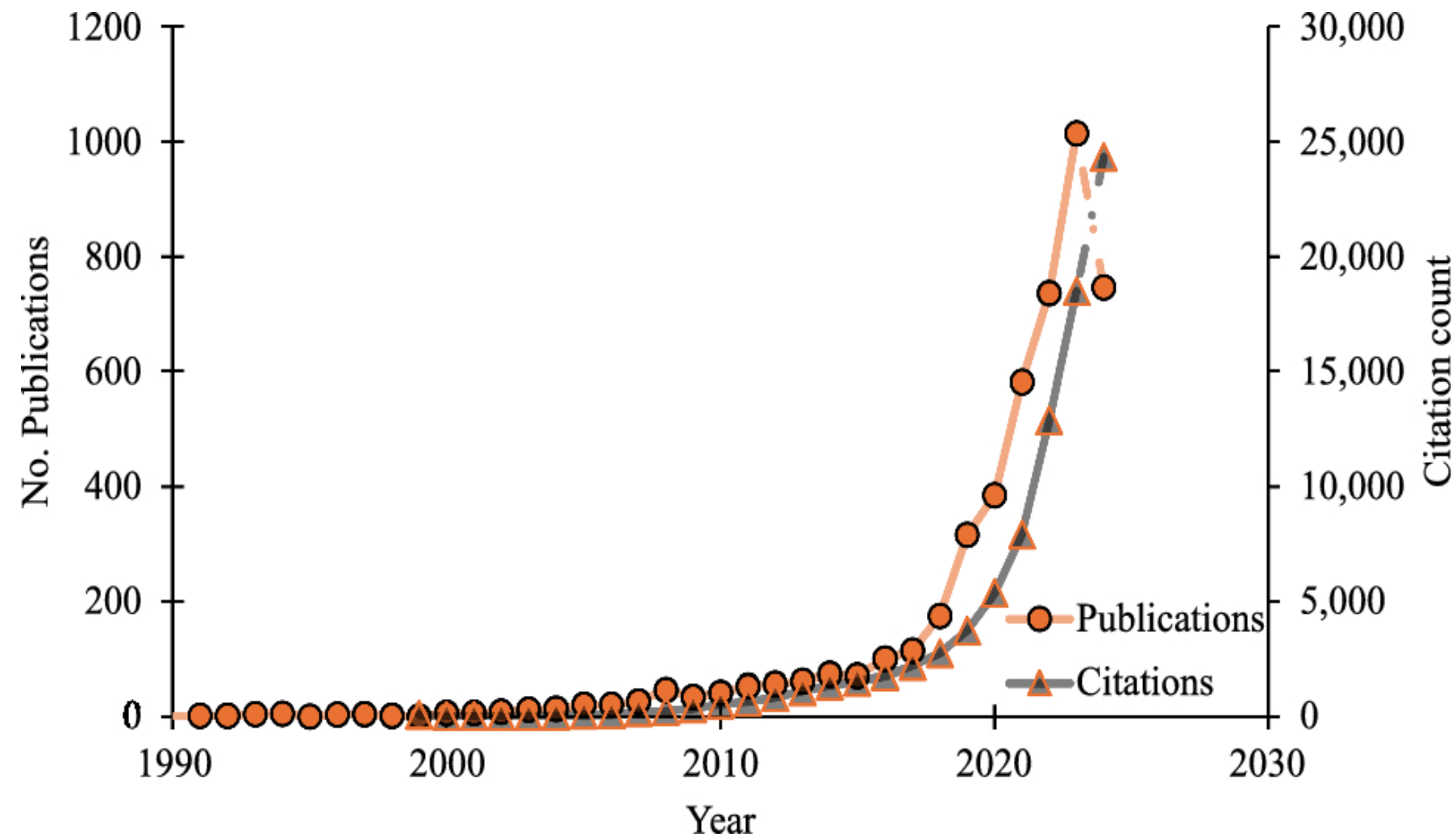


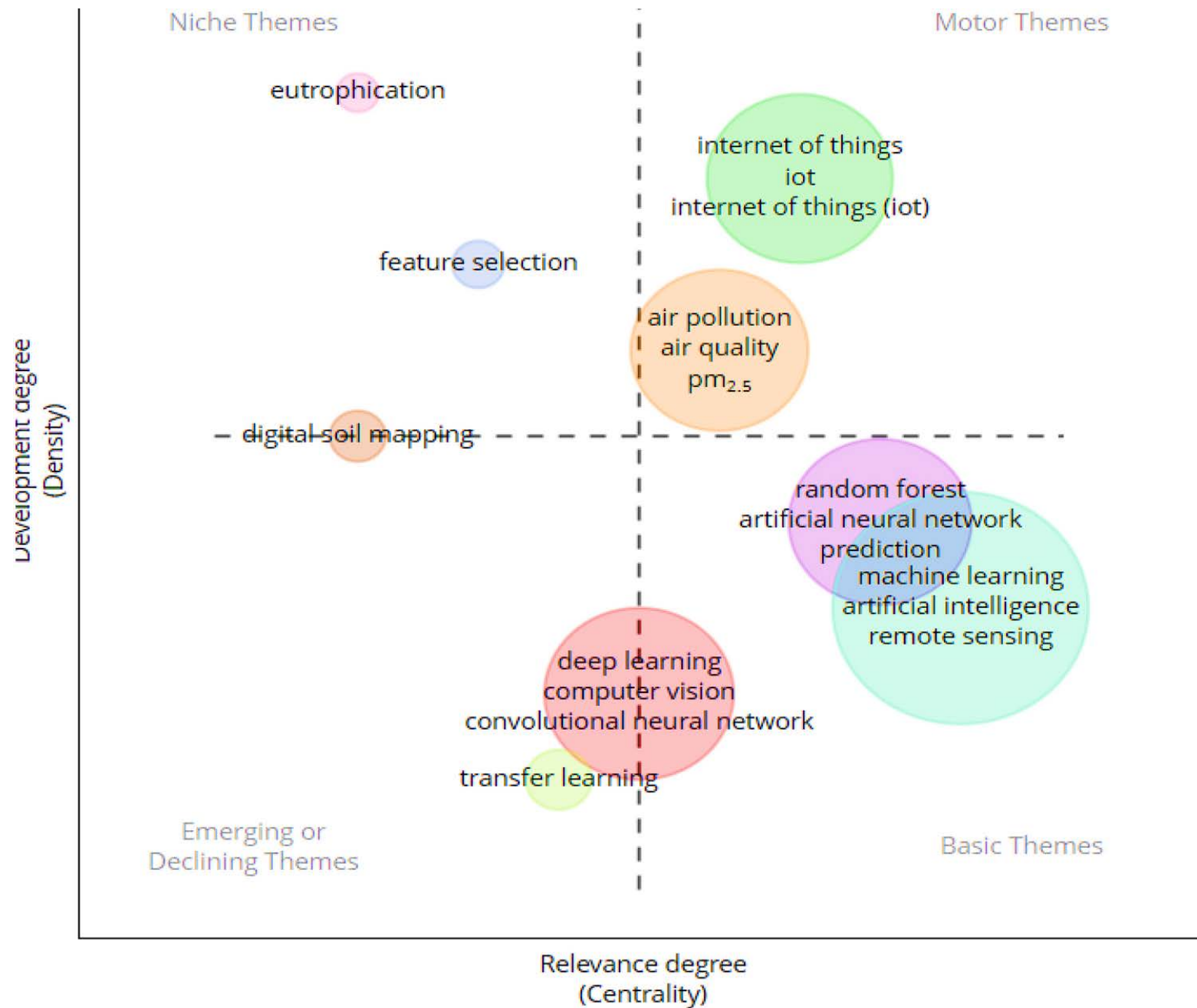


Data-rich but information-poor?



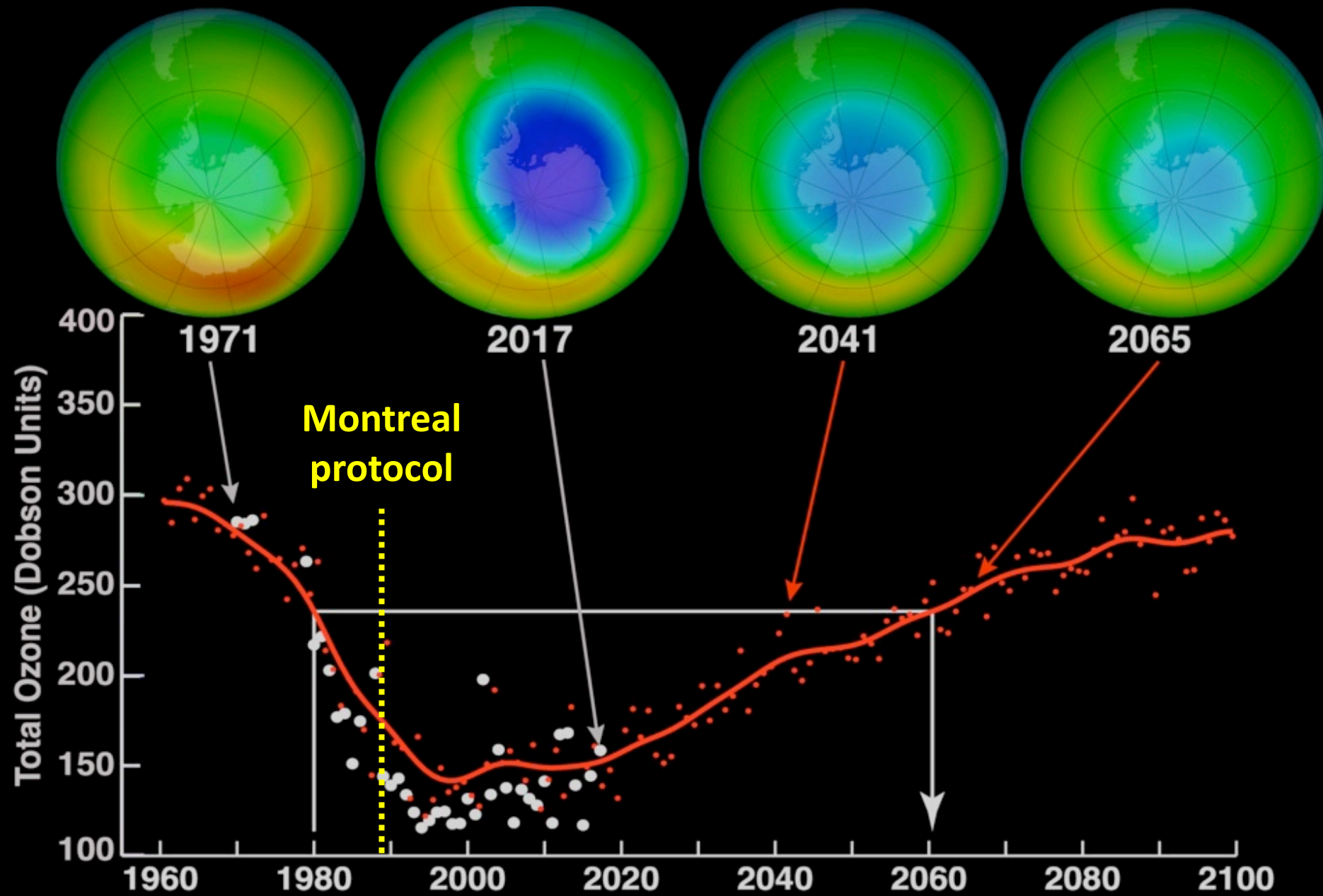
Publications in AI-based Environmental Data Analysis

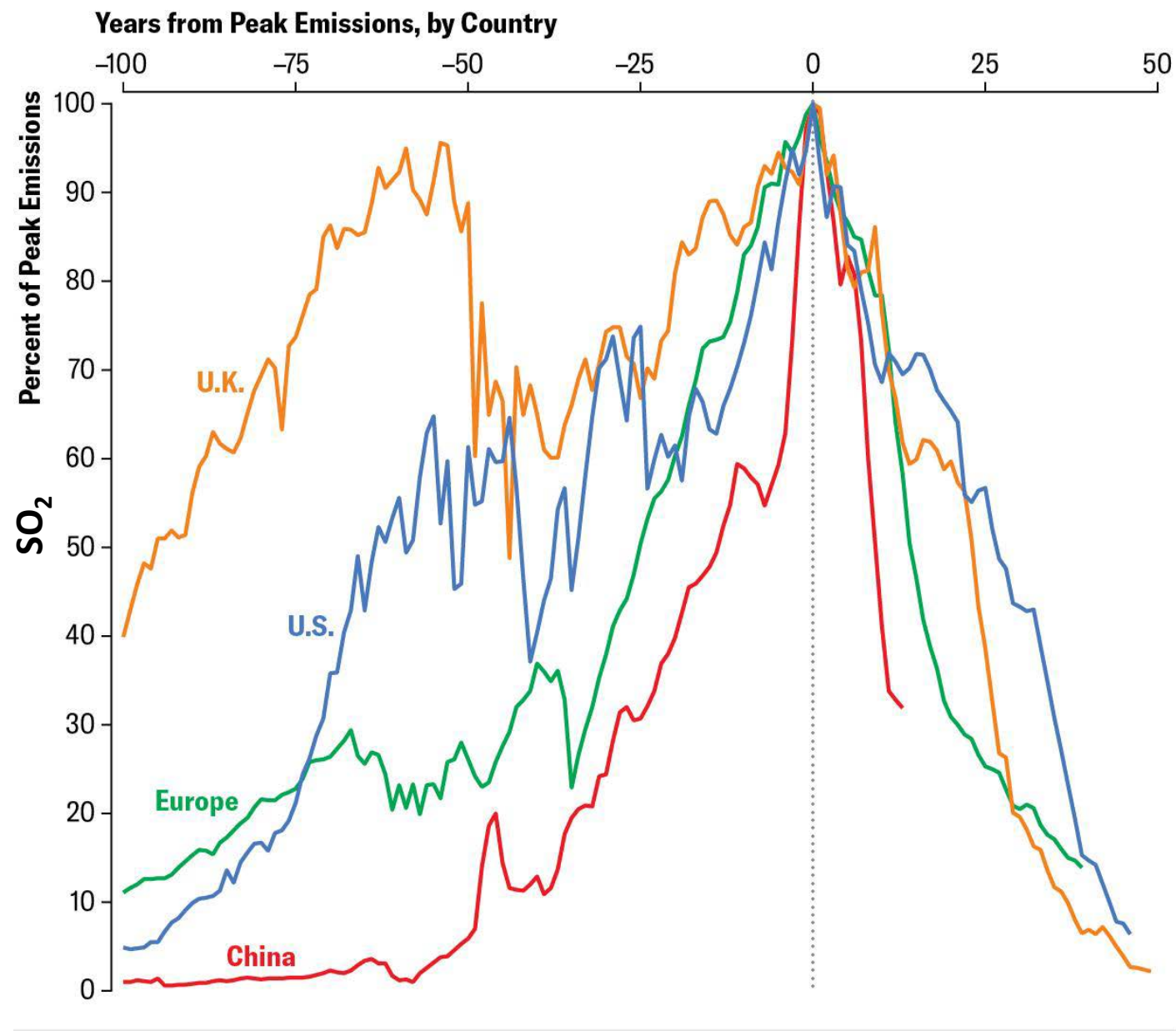






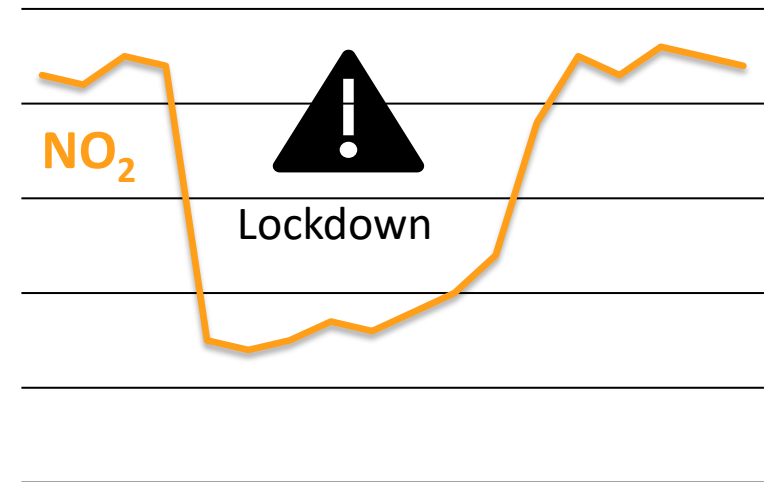
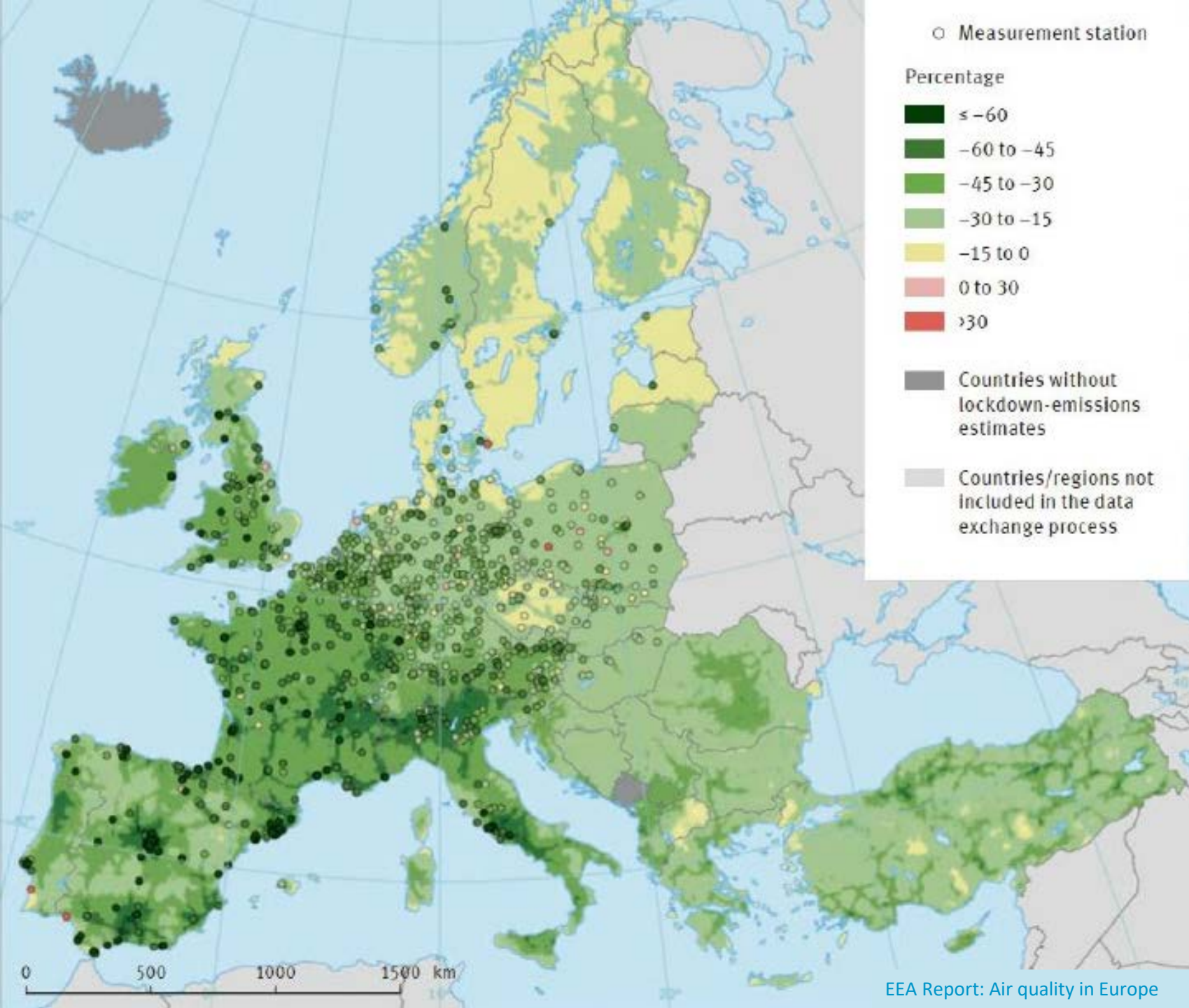
Emerging AI methods can
unlock new data sources.







Environmental policies
do have an effect.





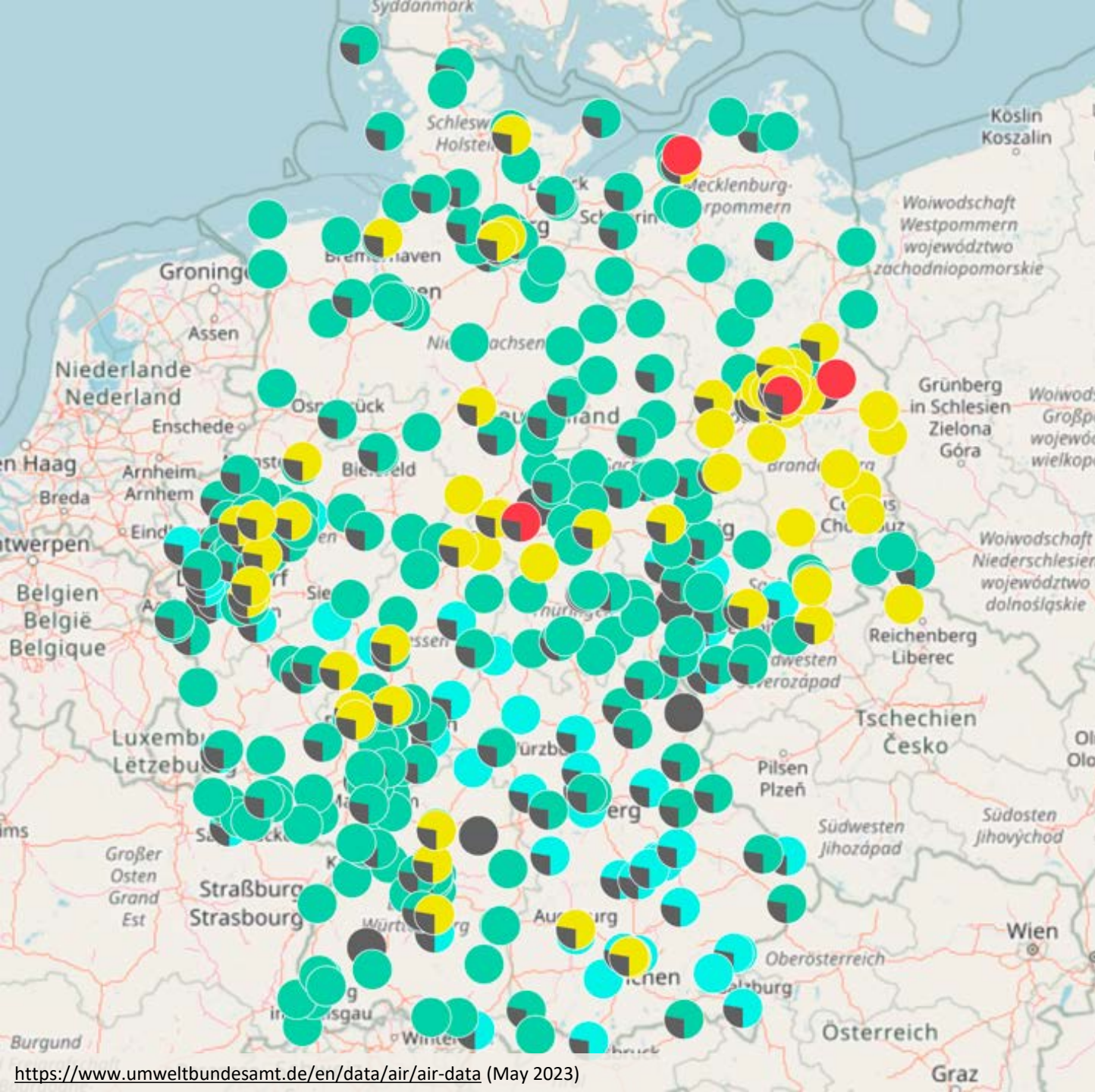
<https://www.bbc.com/news/health-52183295>



Axel Eggert / Umweltbundesamt



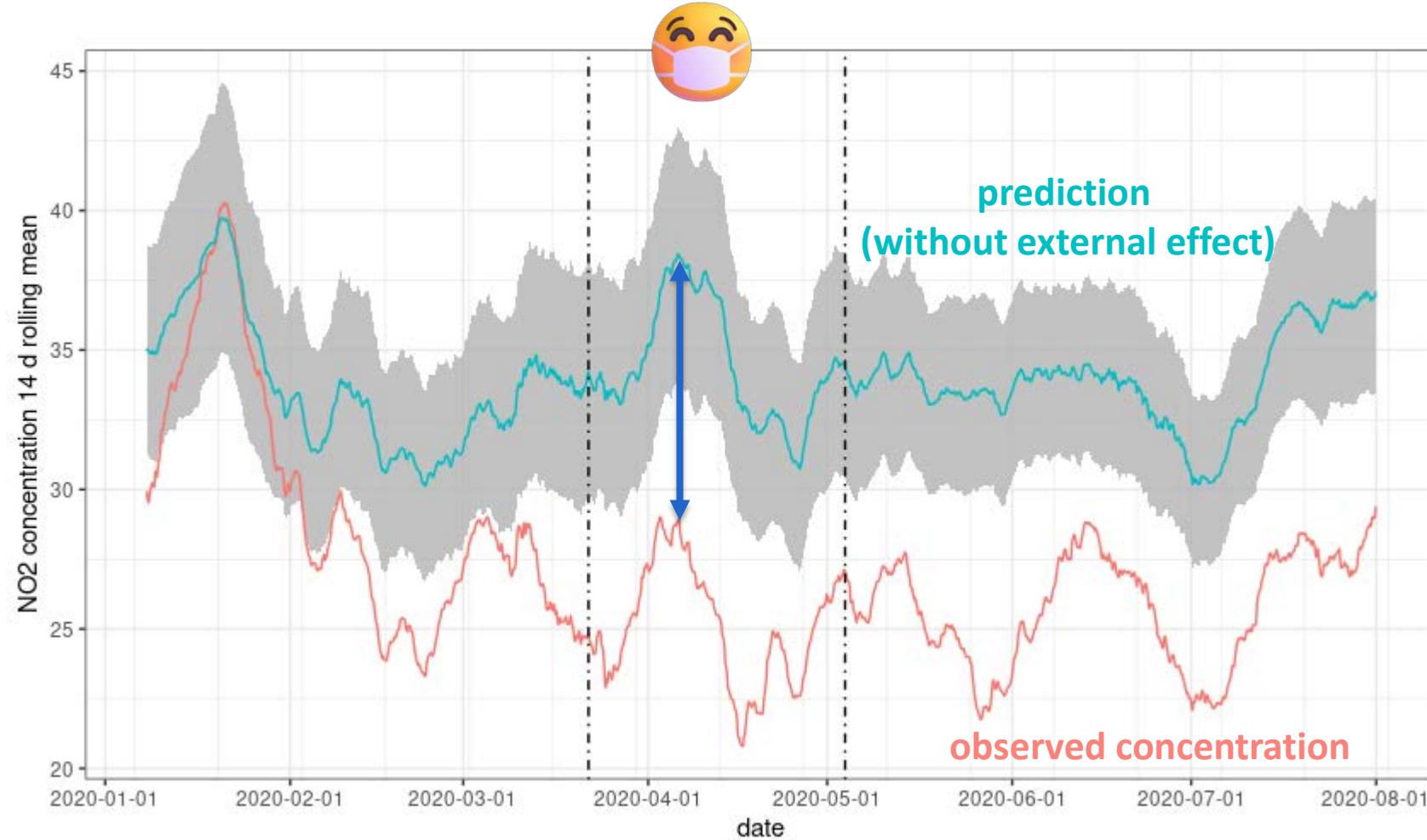
<https://www.nachrichtenleicht.de/neun-euro-ticket-bundestag-100.html>



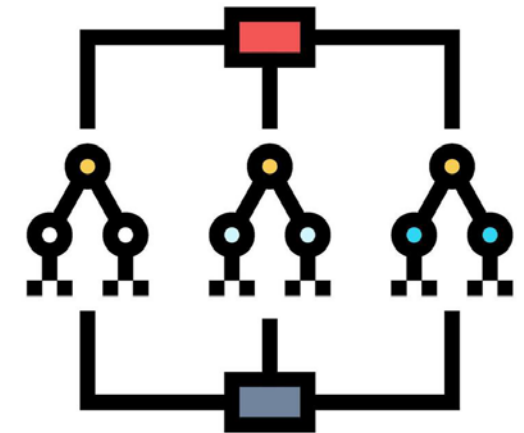
Influences on air quality:

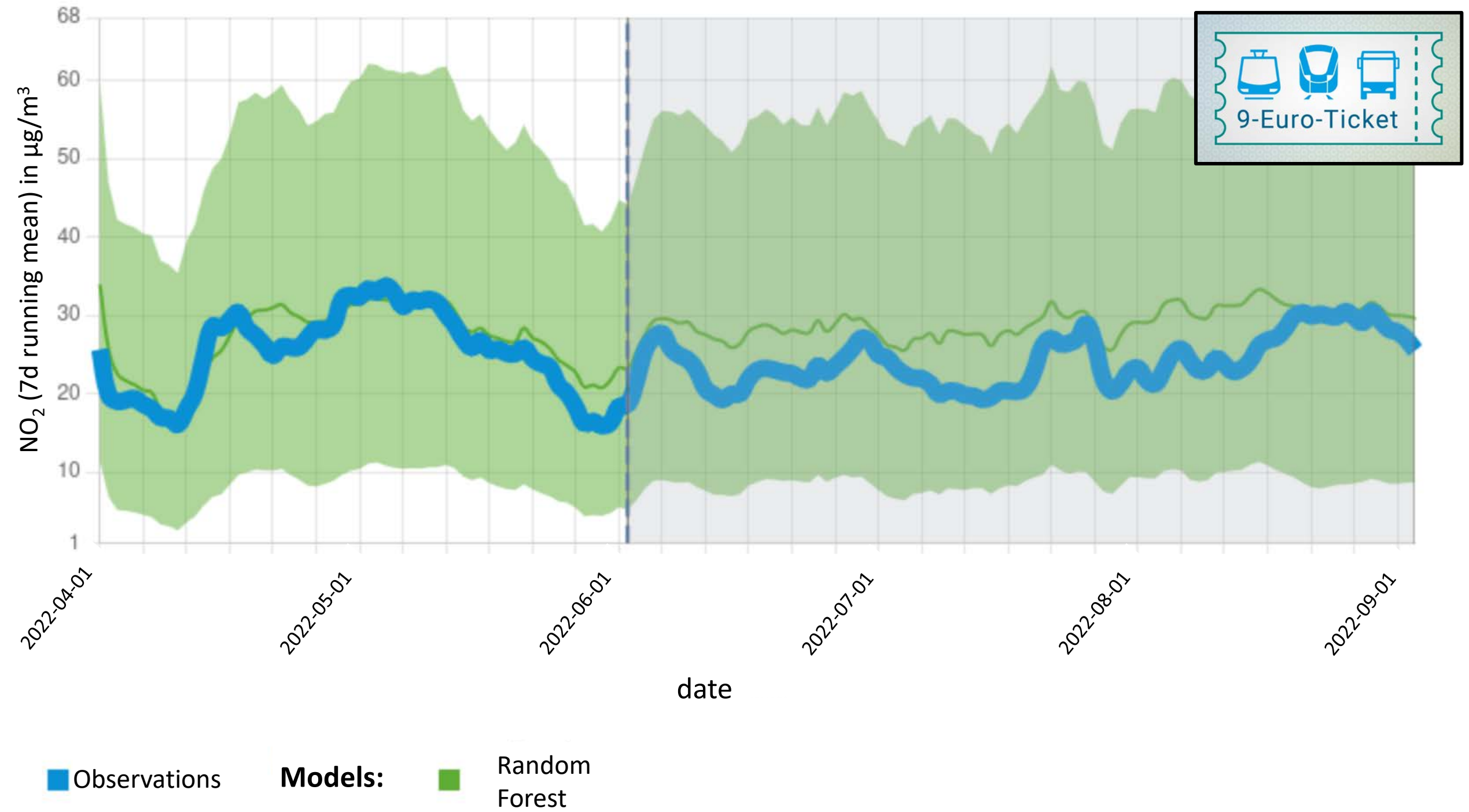
- background concentration
- chemical processes
- weather conditions
- ...

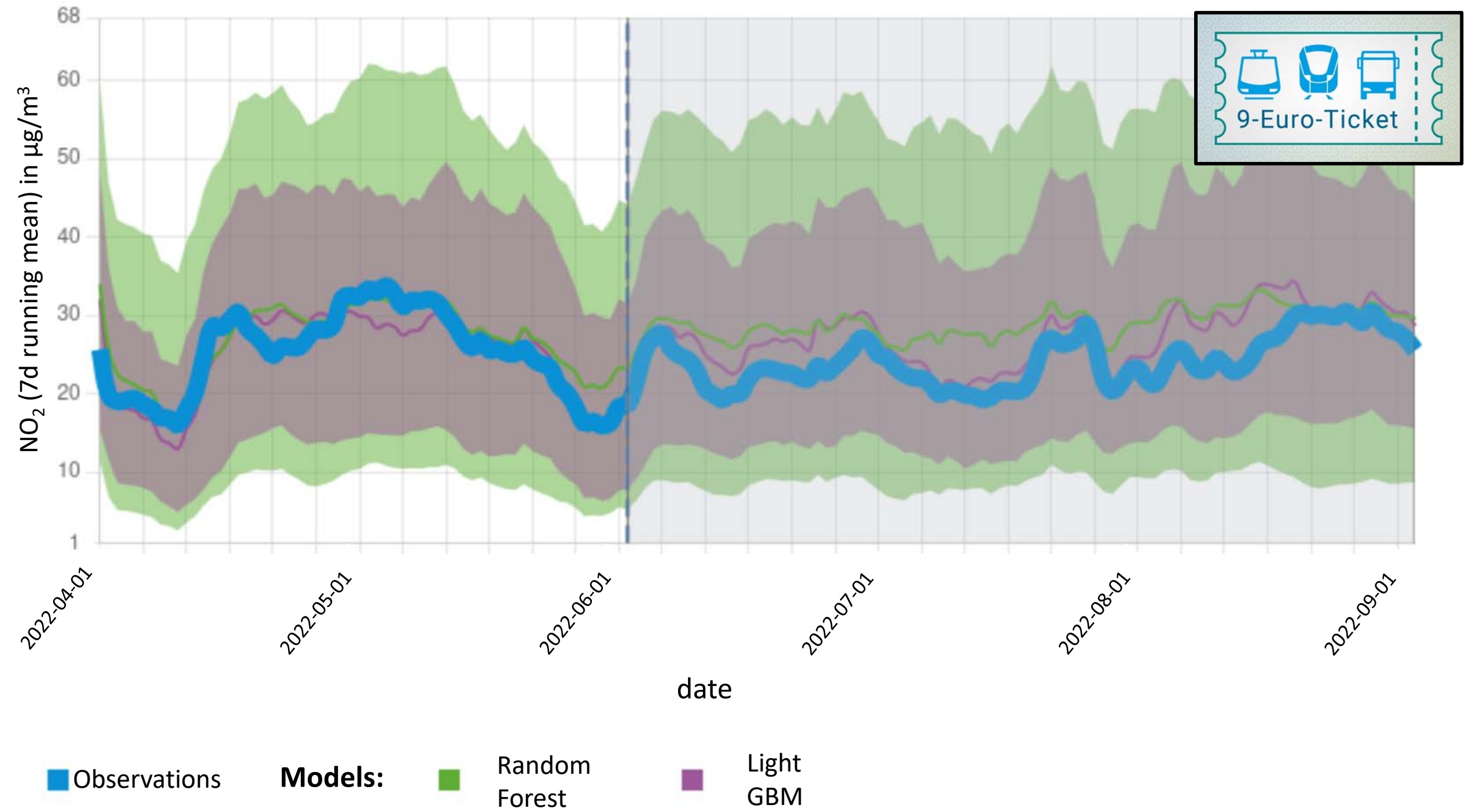
Prediction of air quality during the covid pandemic

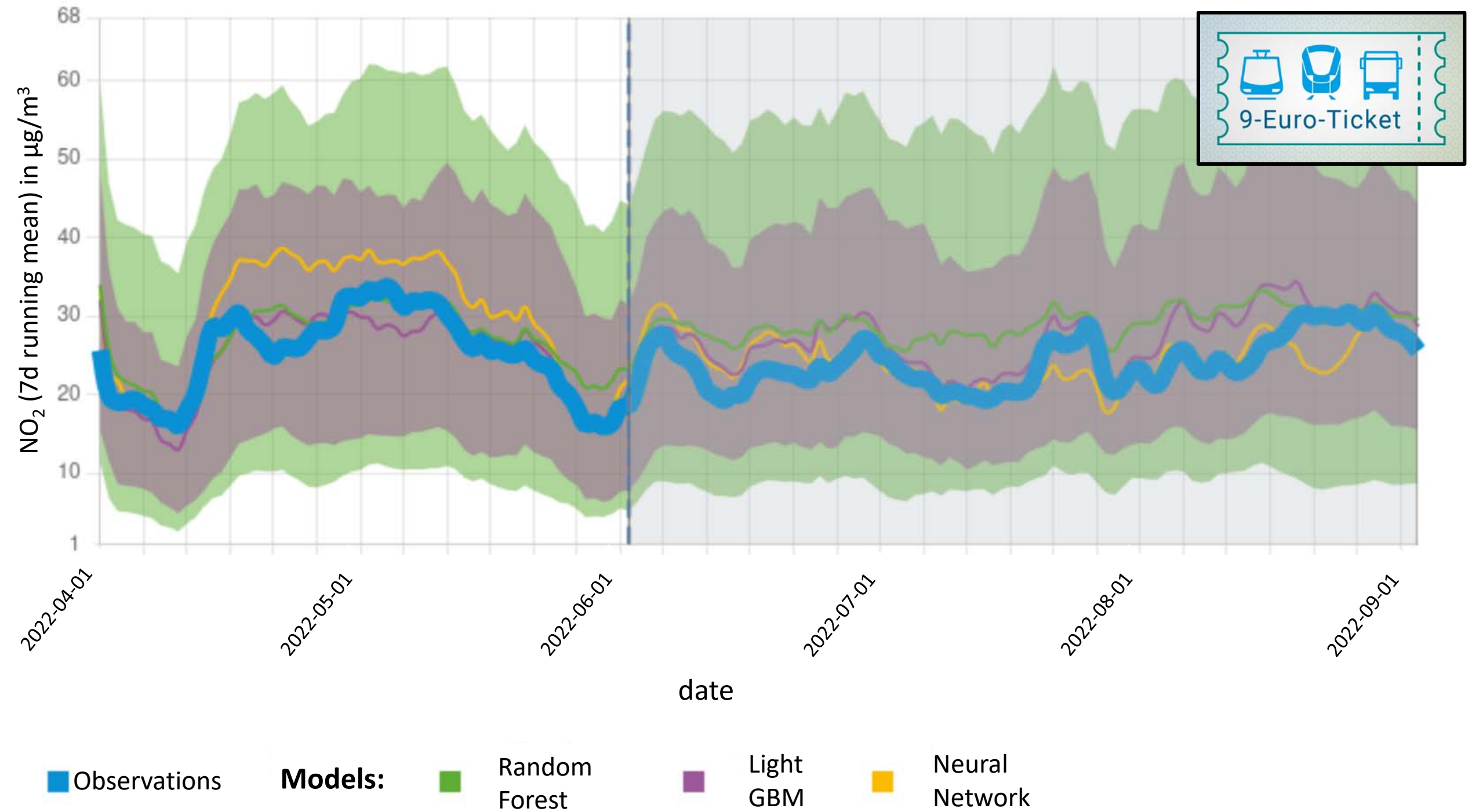


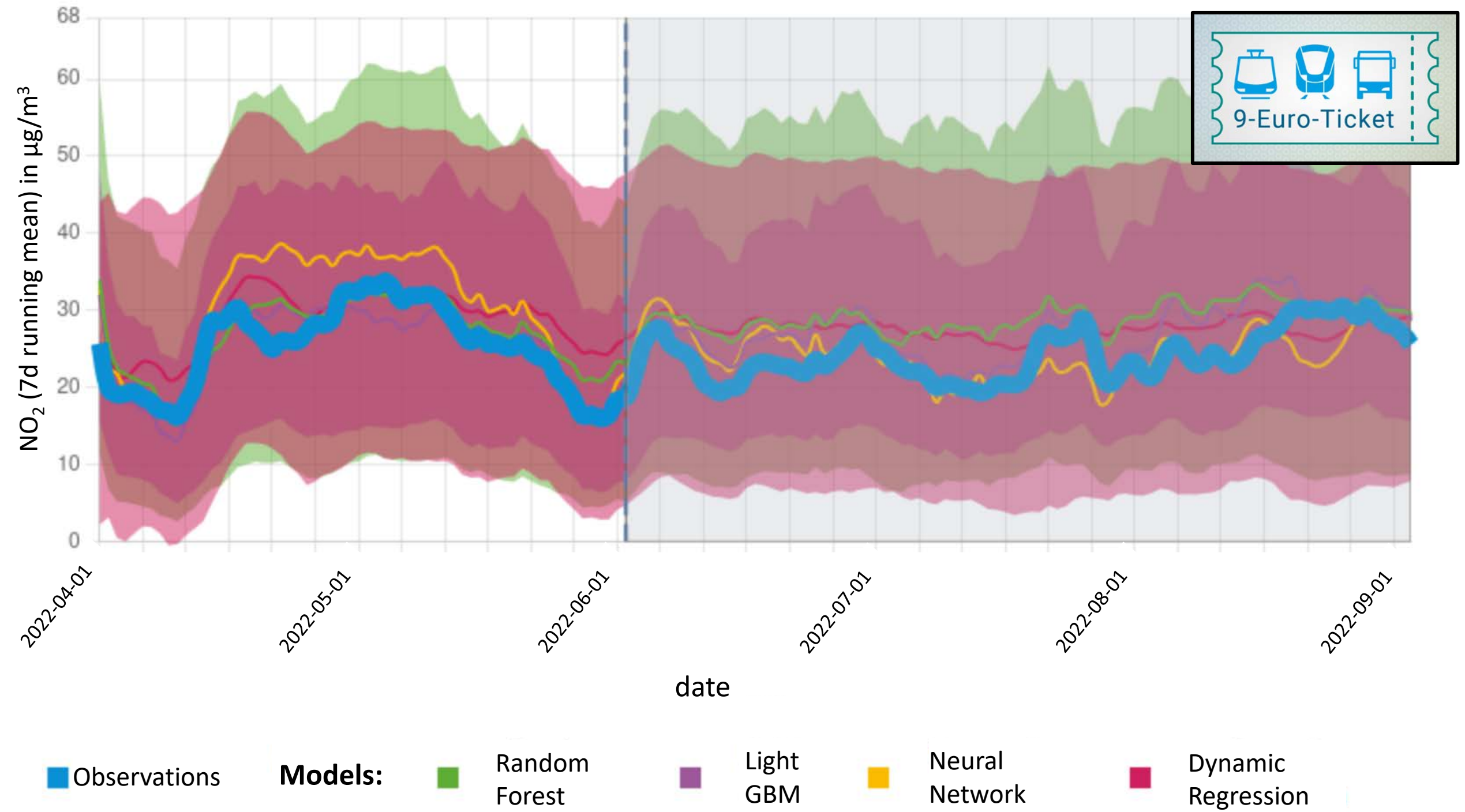
Model:
Random forest





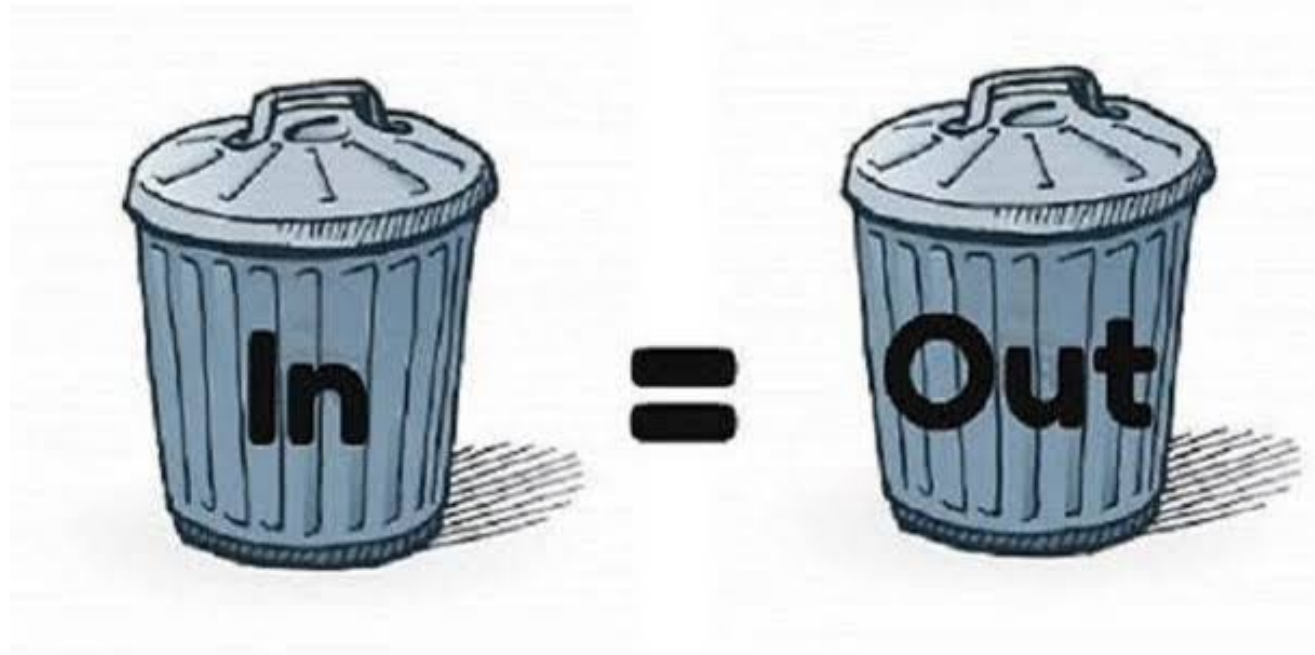








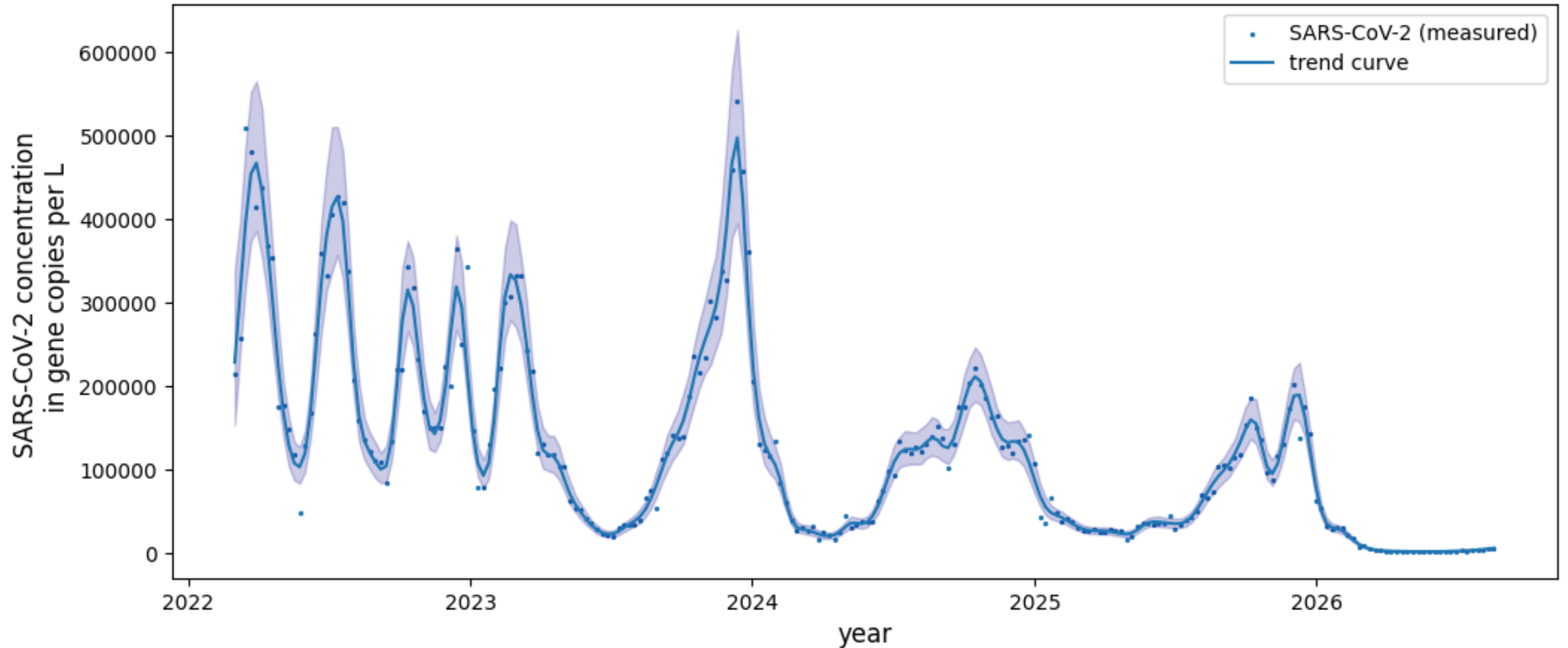
Data Science helps to quantify
effects of political decisions.

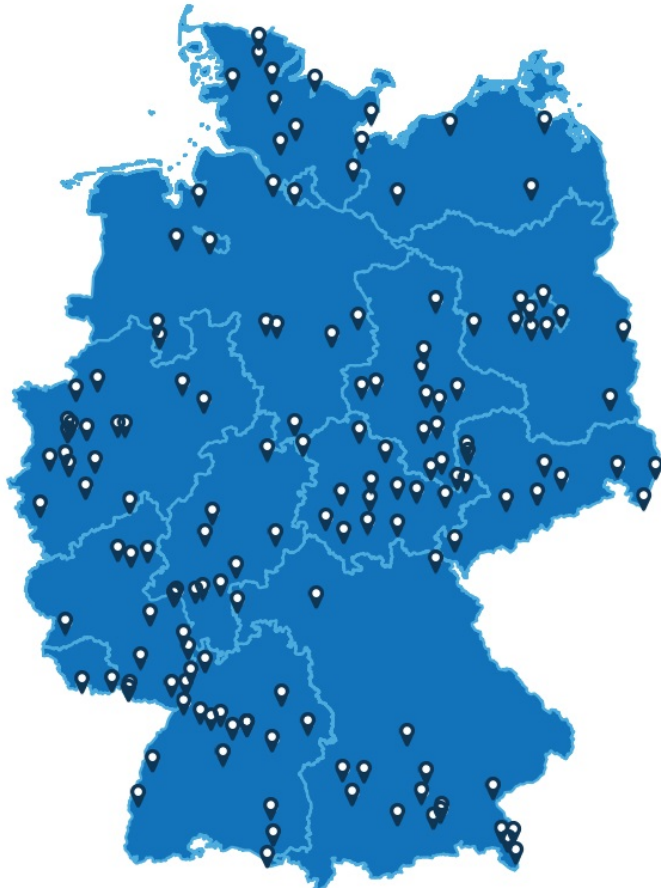




Data quality assessment in wastewater surveillance

Wastewater-based Epidemiology



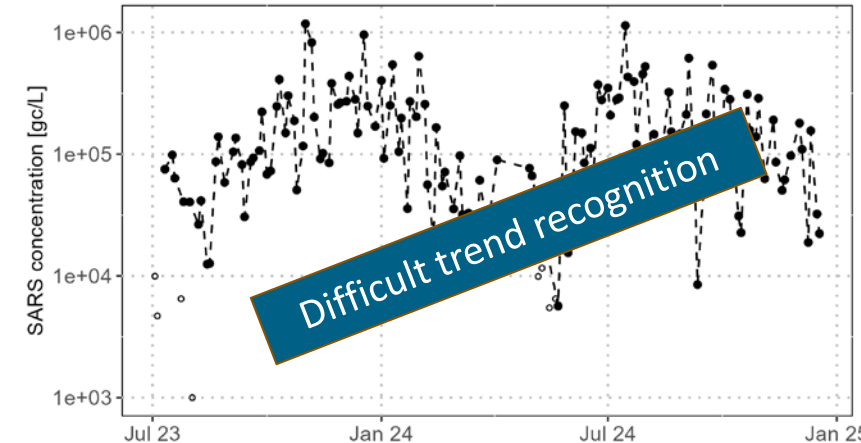
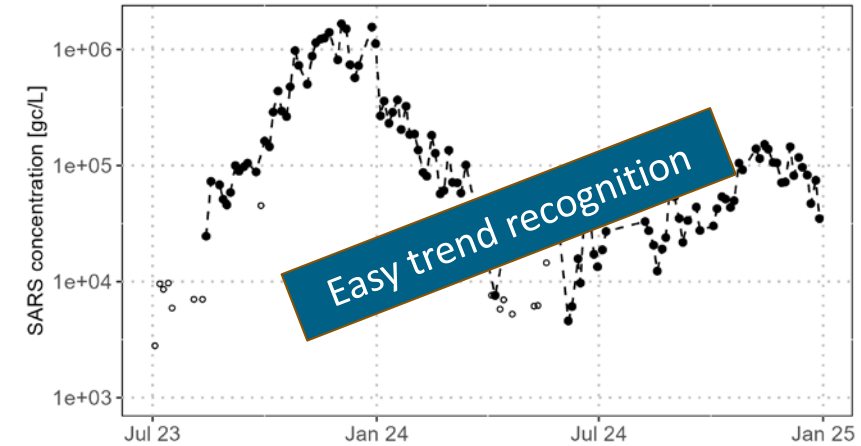


169 participating WWTPs
2022-2024

Challenges

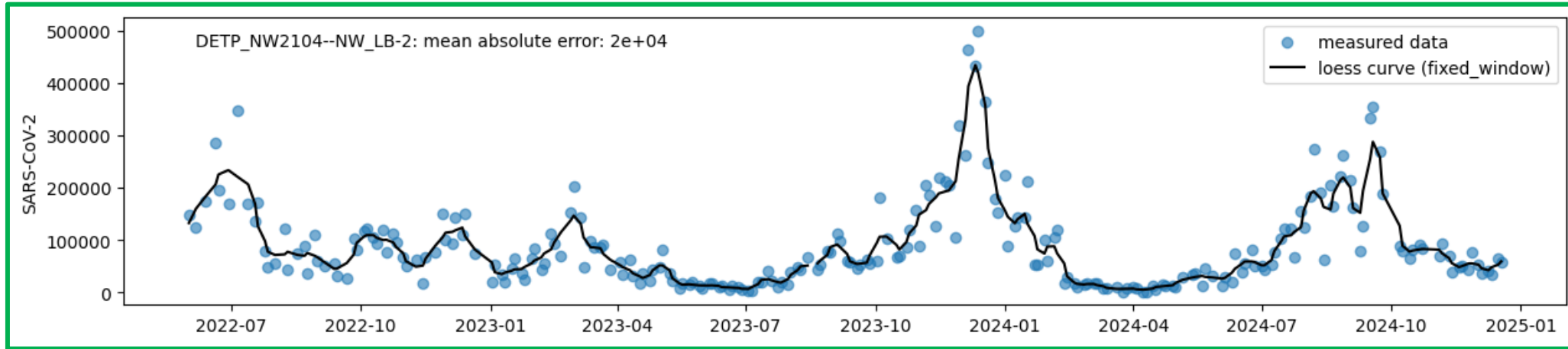
- operational costs
- inconsistent data quality between different wastewater treatment plants / laboratories

→ Which locations are really worth sampling?

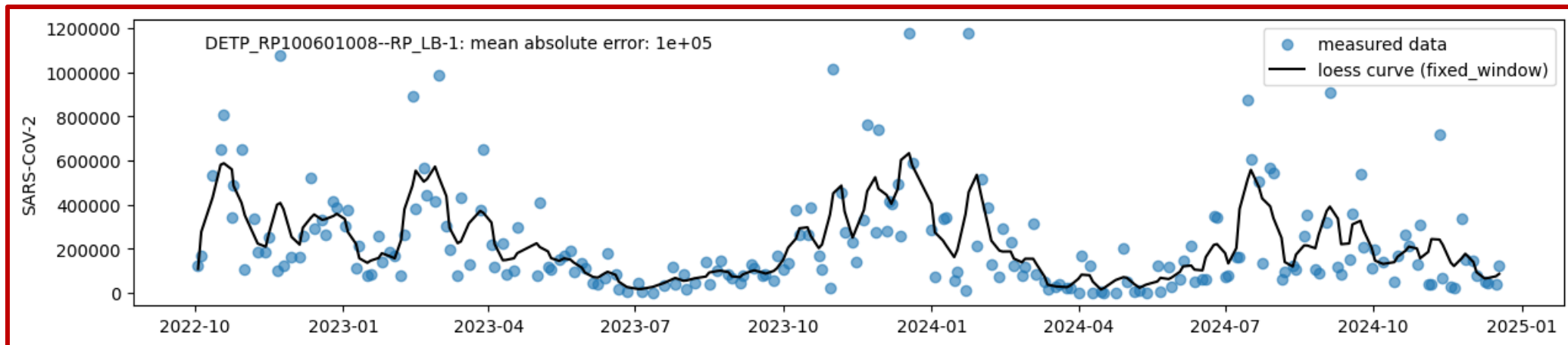


Mean absolute error from LOESS* curve

*Locally Estimated
Scatterplot Smoothing

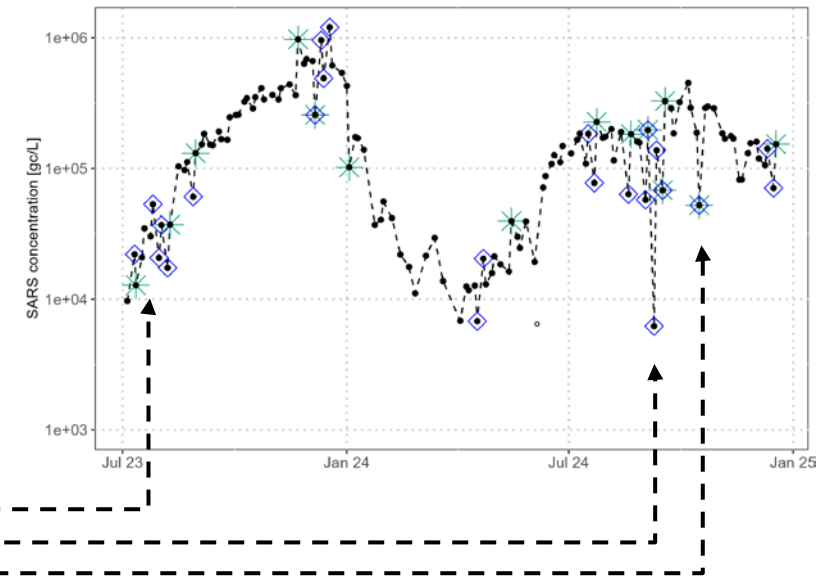


low MAE

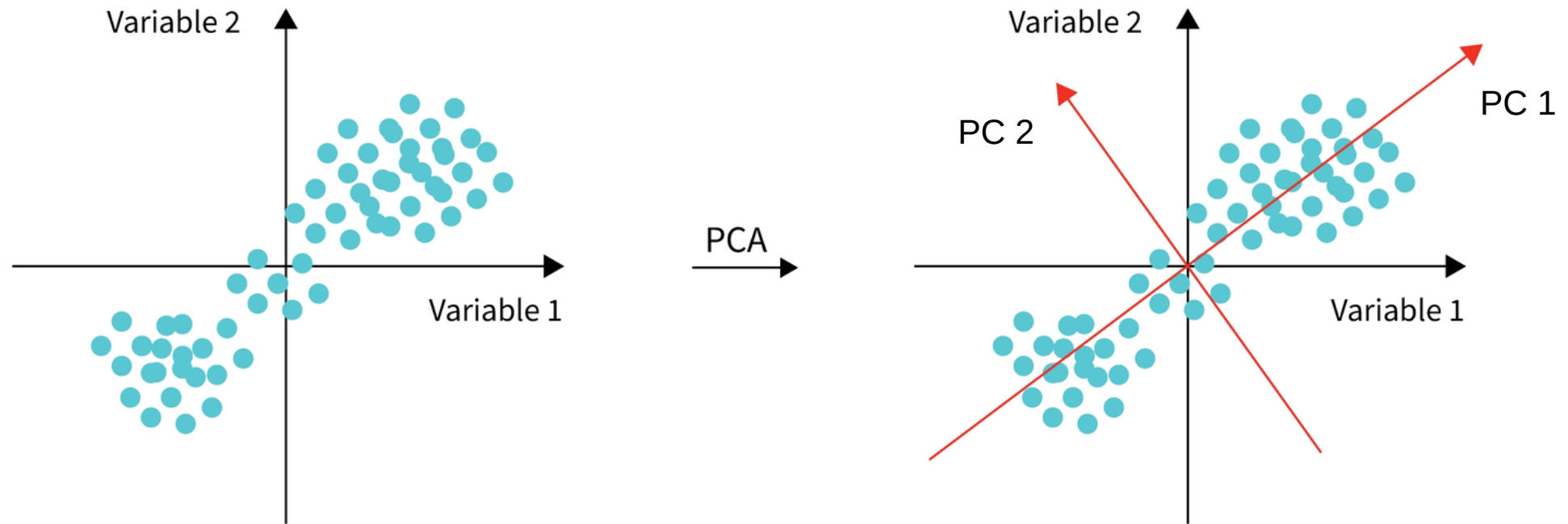


high MAE

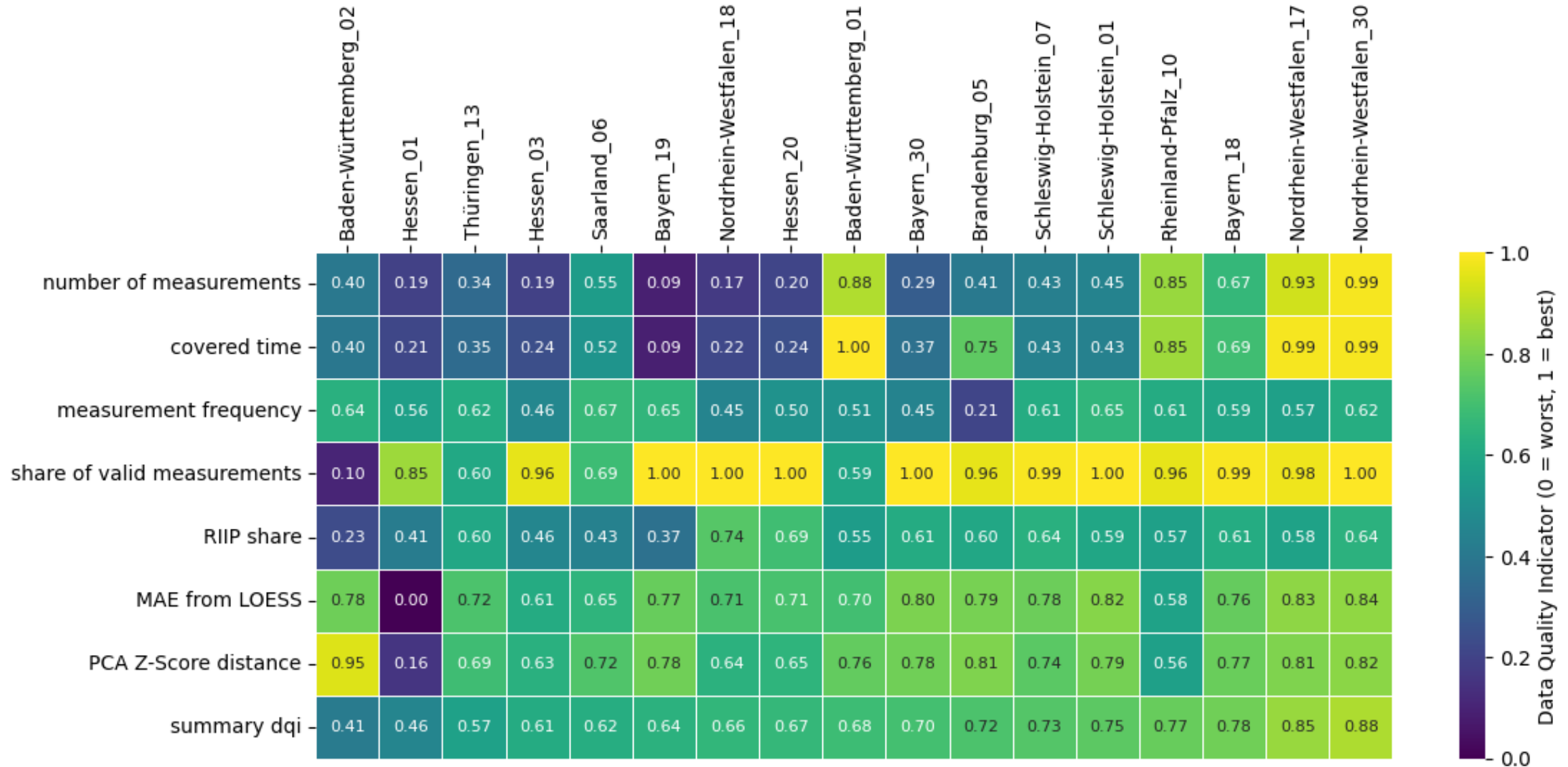
Principal Component Analysis (PCA)

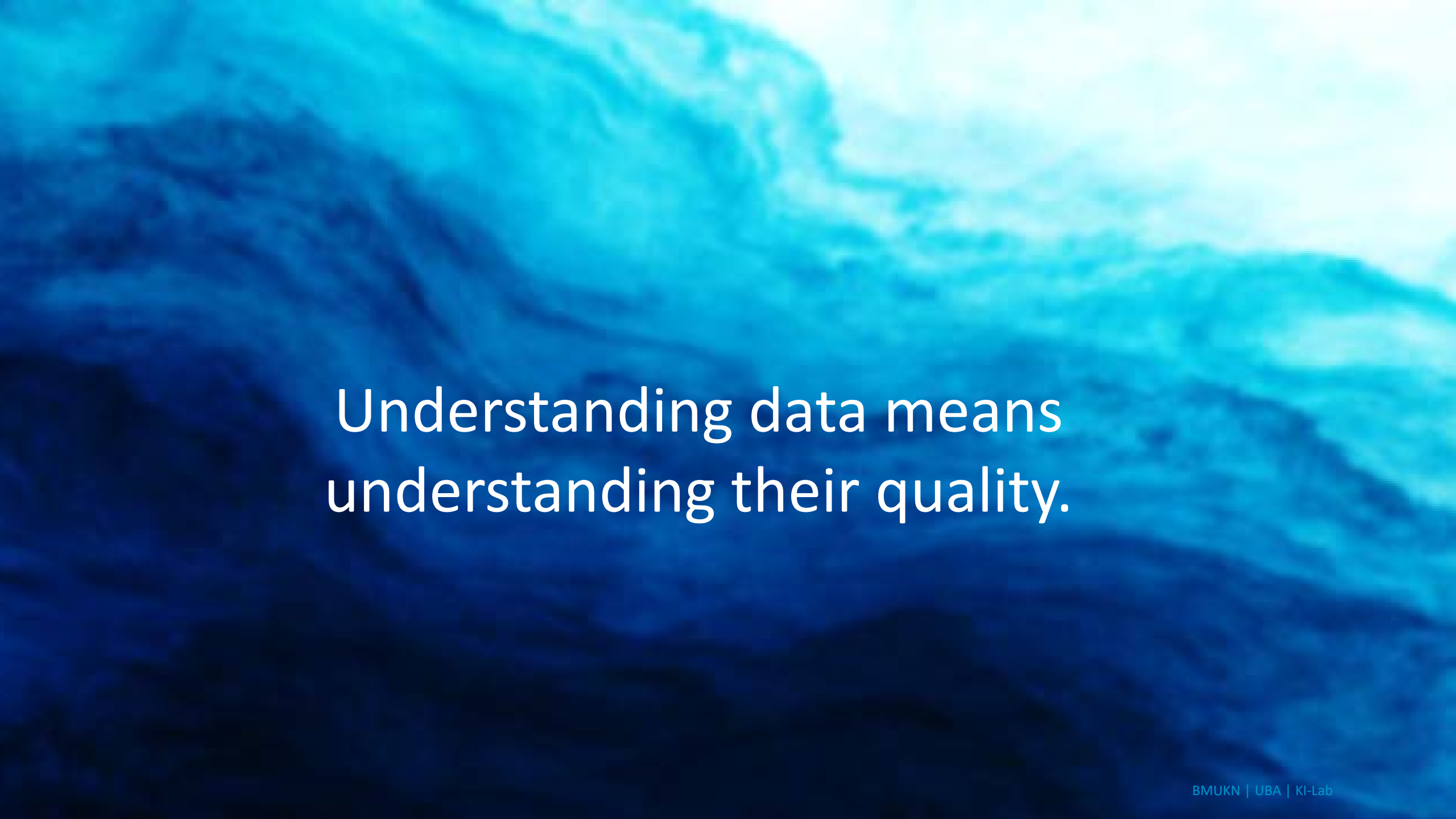


Principal Component Analysis (PCA)



Results: heatmap of Data Quality Indicators





Understanding data means
understanding their quality.



Data alone will not protect the environment.

Understanding them will.
may.



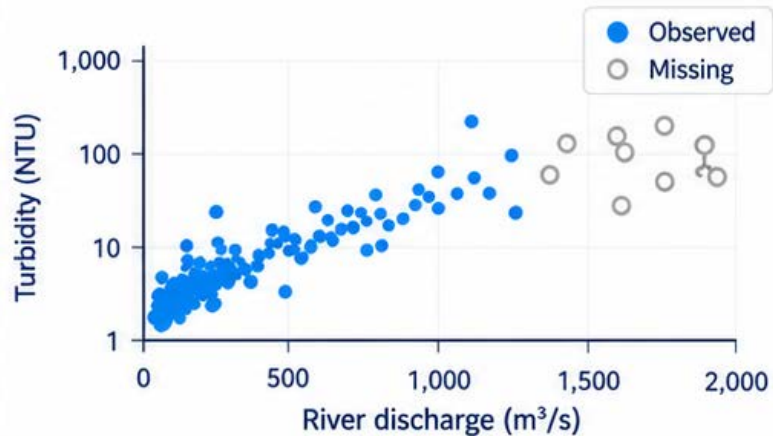
MCAR

(Missing Completely At Random)



MAR

(Missing At Random)



MNAR

(Missing Not At Random)



	A	B	C
0			
1			
2			
3			
4			

data imputation

single imputation

fill gaps with
other value

univariate

multivariate

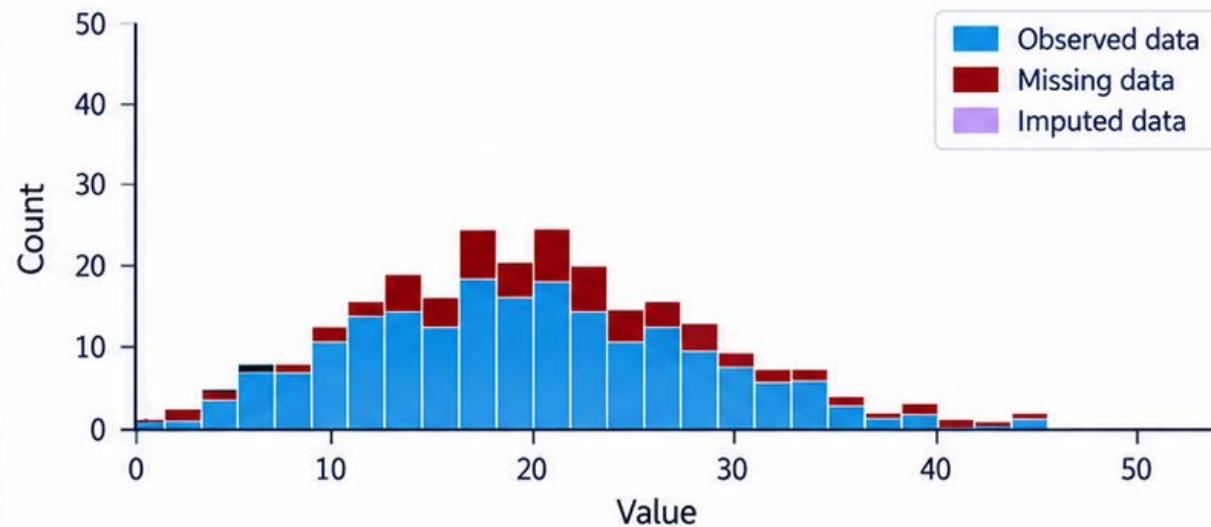
data dropping

delete incomplete
rows

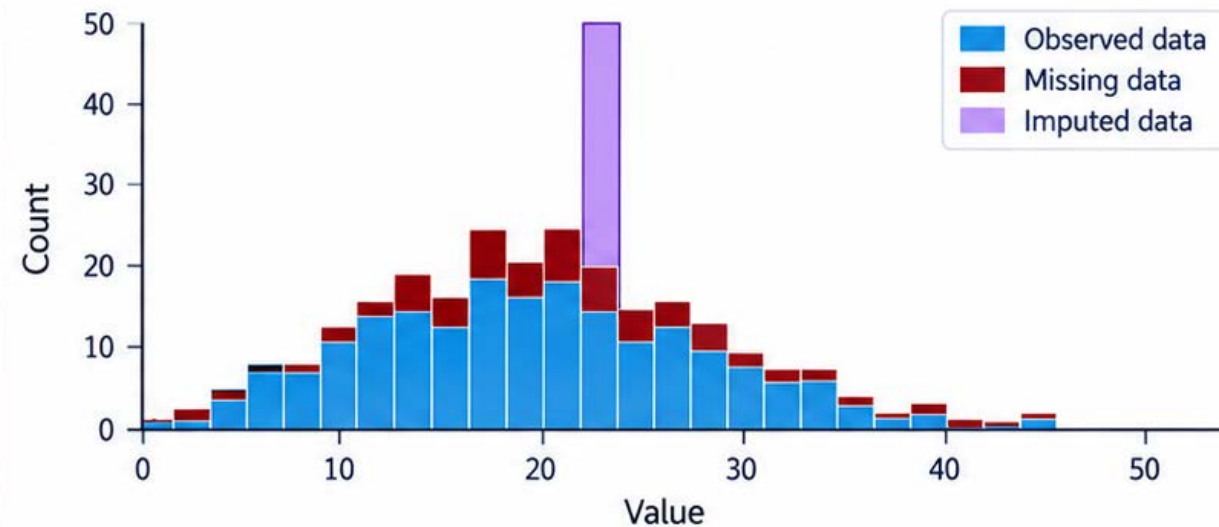
multiple imputation

fill gaps &
include uncertainties

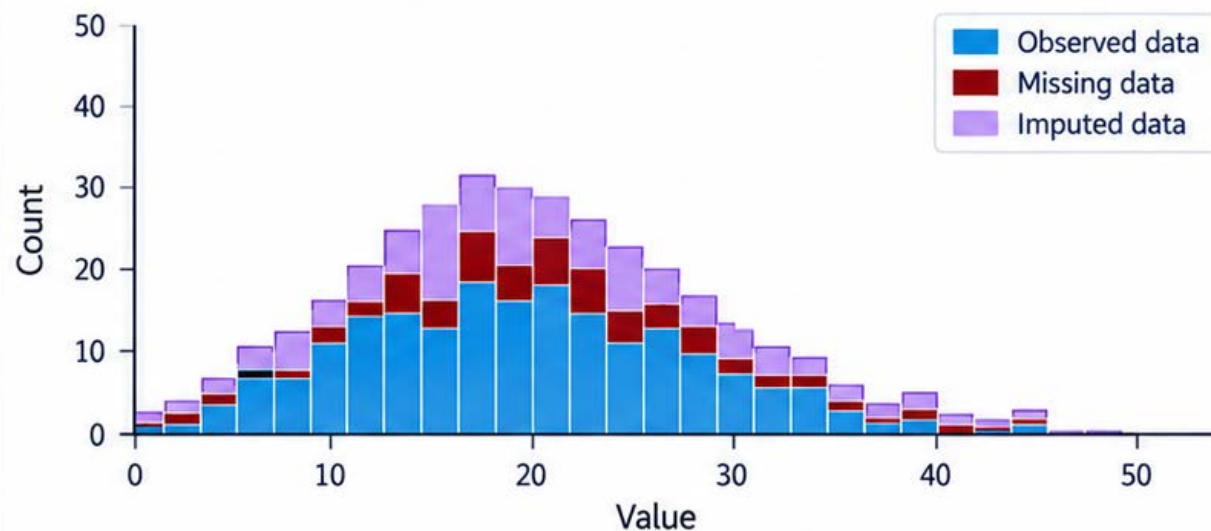
Dropping



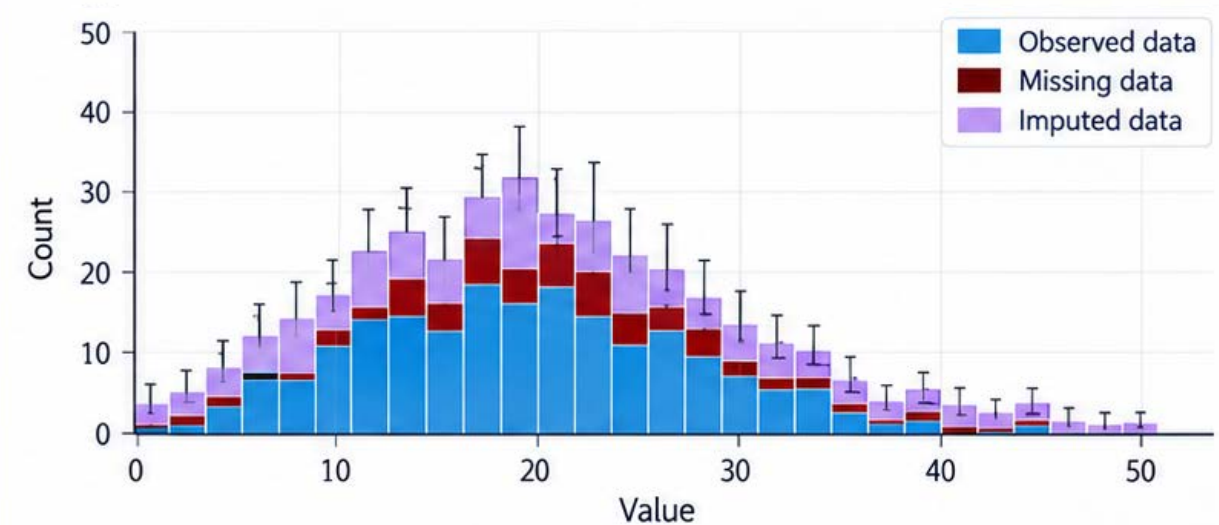
Mean imputation



k-nearest-neighbours imputation



Bootstrap imputation





Understand your data gaps
to understand your data.