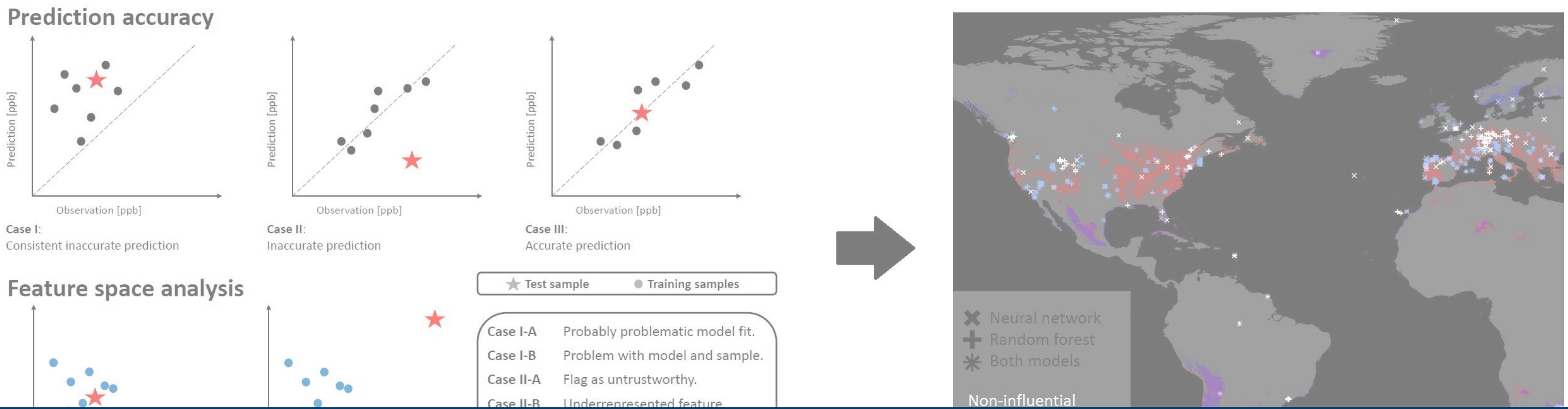


Latent space, feature space and the global domain – how ozone research can benefit from explainable machine learning

ECMWF Machine Learning Workshop

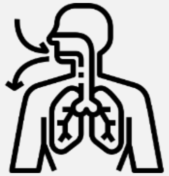
31 MAR 2022 I CLARA BETANCOURT | c.betancourt@fz-juelich.de

Member of the Helmholtz Association

JÜLICH Forschungszentrum

TROPOSPHERIC OZONE

A toxic greenhouse gas



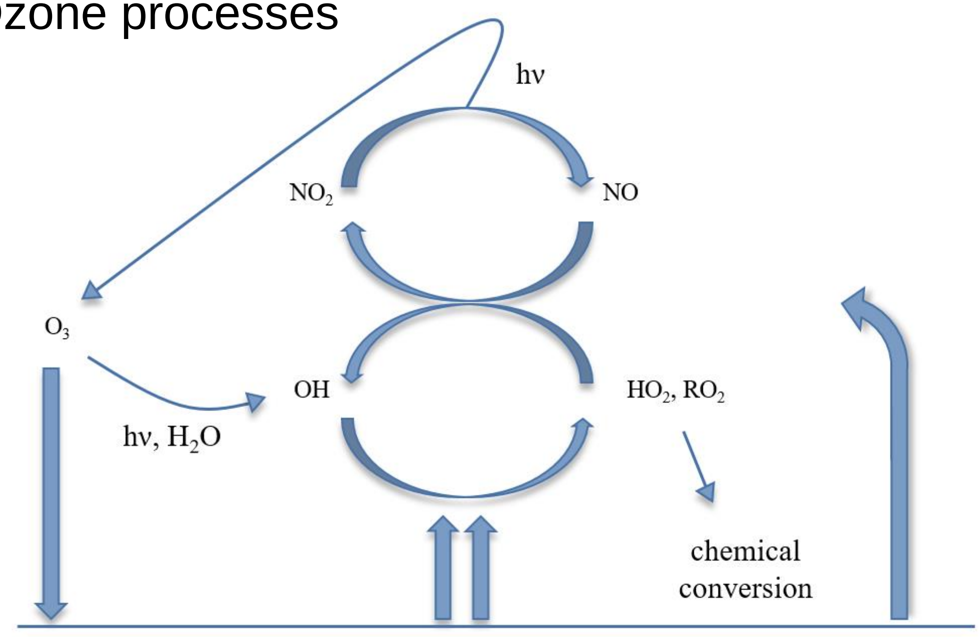
Lung damage
→ Global Burden of Disease Study 2017

Vegetation damage
→ Emberson et al, 2018

Short lived climate forcer
→ IPCC, 2018

Icons by Flaticon

Ozone processes



→ Betancourt et al. 2021

c.betancourt@fz-juelich.de

JÜLICH Forschungszentrum

OUTLINE

Explainable machine learning for ozone research

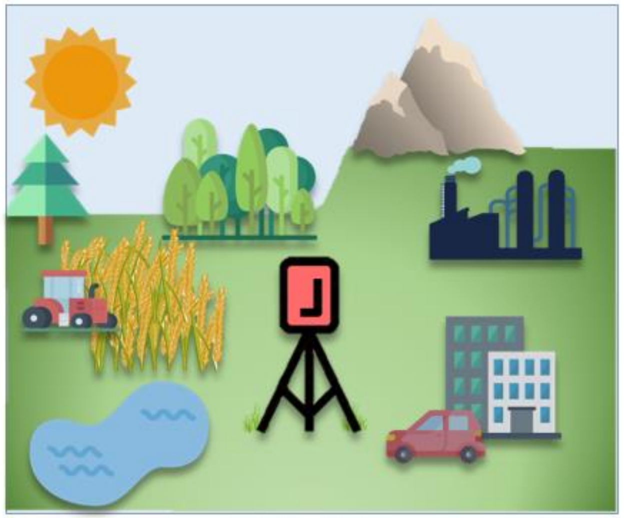
- AQ-Bench dataset
- Explaining and improving faulty predictions
- Ozone mapping on the global domain

c.betancourt@fz-juelich.de

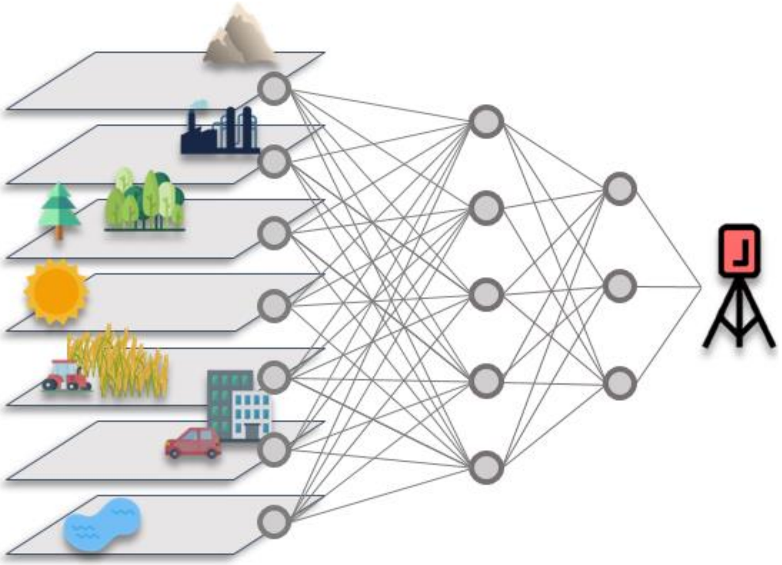
AQ-BENCH DATASET

A benchmark dataset for machine learning on global ozone metrics

The AQ-Bench dataset contains long-term air quality metrics and metadata at sites around the globe.



The air quality at a site is influenced by its surroundings.



The proposed machine learning task is to train a machine learning algorithm which maps from metadata to long-term air quality metrics at measurement sites.

→ Betancourt et al. 2021

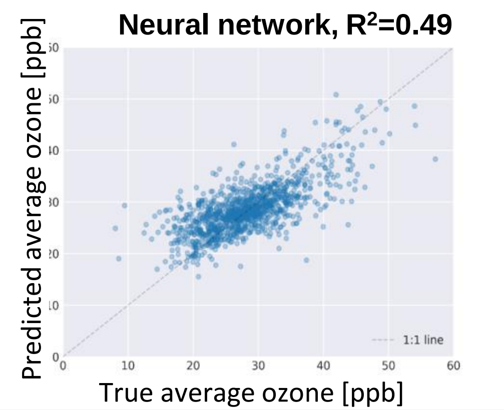
OUTLINE

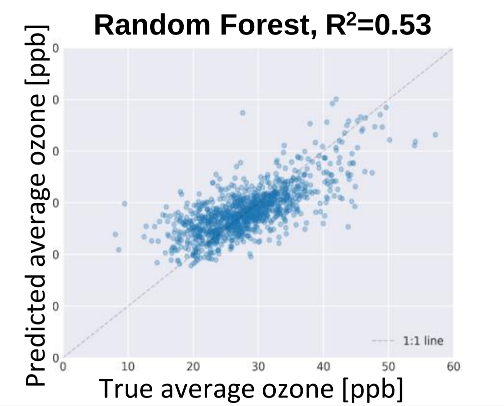
Explainable machine learning for ozone research

- AQ-Bench dataset
- Explaining and improving faulty predictions
- Ozone mapping on the global domain

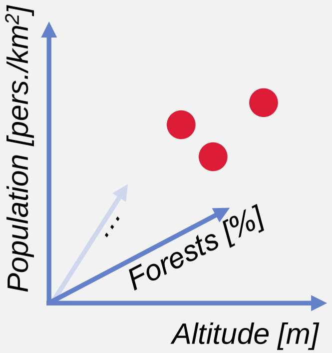
NEURAL NETWORK AND RANDOM FOREST

Test evaluation metrics are fine, but why are some test predictions biased?

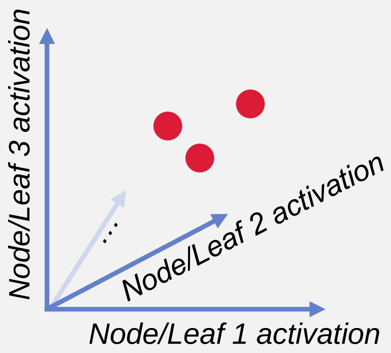




Feature space



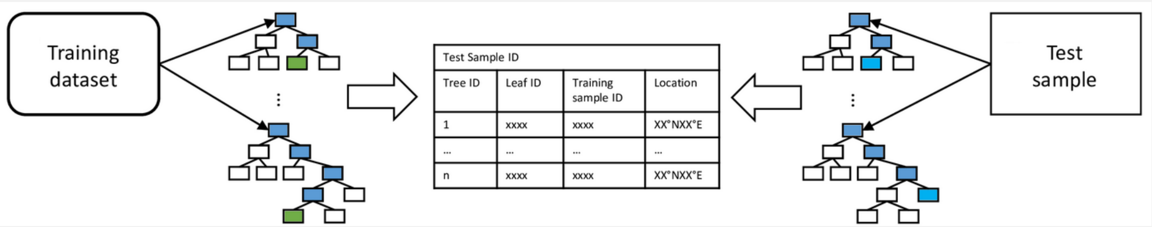
Latent space



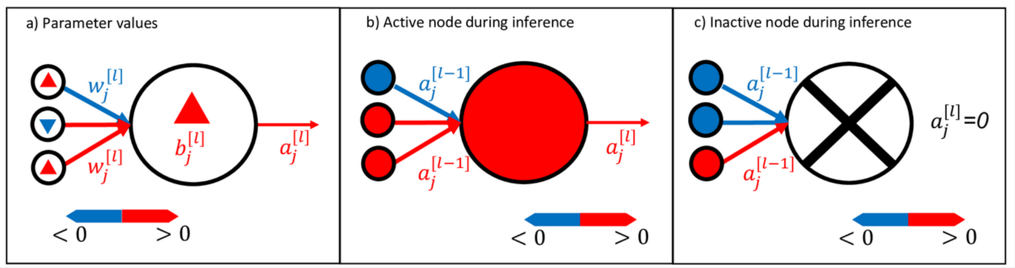
Geographical space

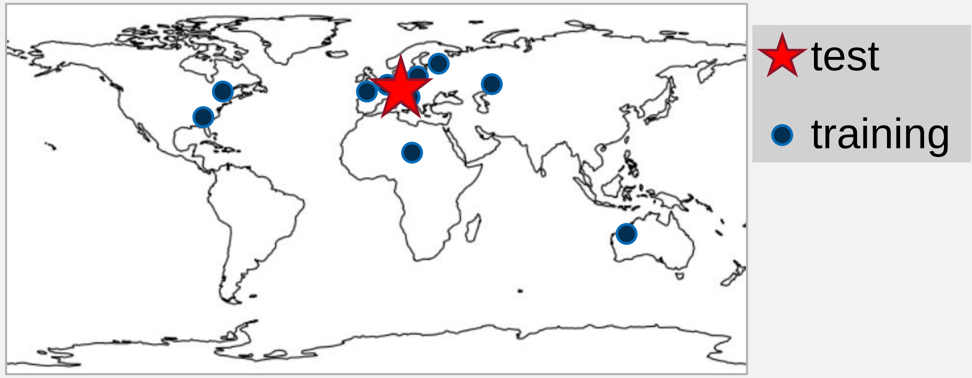
NEAREST NEIGHBORS IN LATENT SPACE

Random forest activations



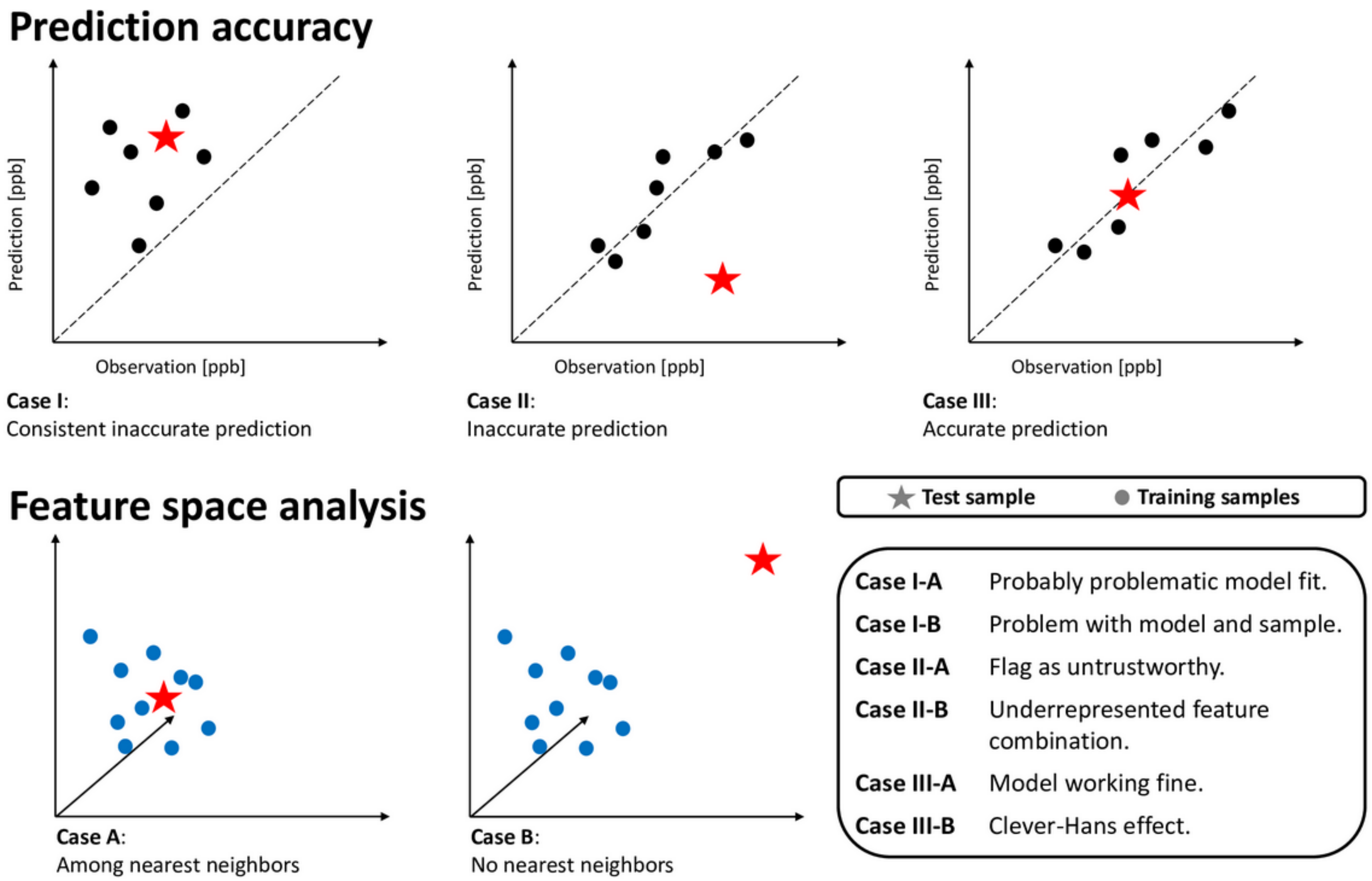
Neural network activations



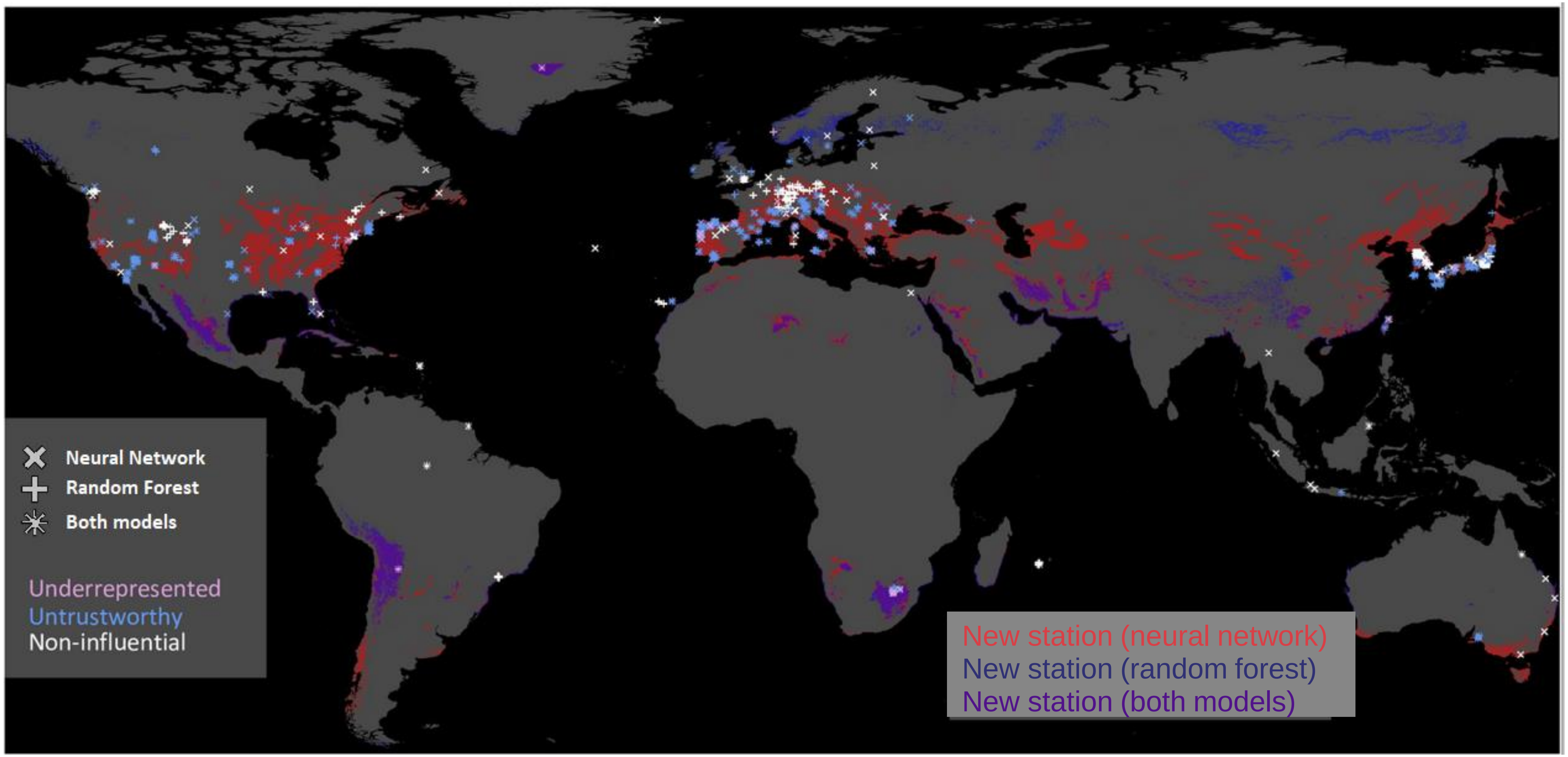


EXPLAINING INACCURATE PREDICTIONS

By combining latent space neighbors and feature space



GLOBAL DOMAIN: WHAT CAN WE LEARN FROM THIS?



→ Stadtler et al. 2021

OUTLINE

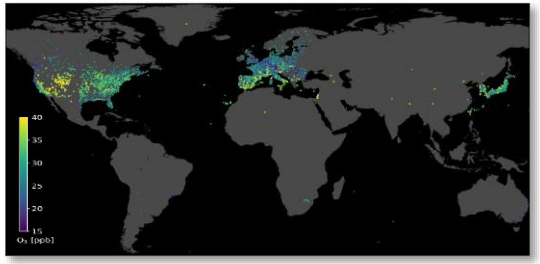
Explainable machine learning for ozone research

- AQ-Bench dataset
- Explaining and improving faulty predictions
- Ozone mapping across the global domain

c.betancourt@fz-juelich.de

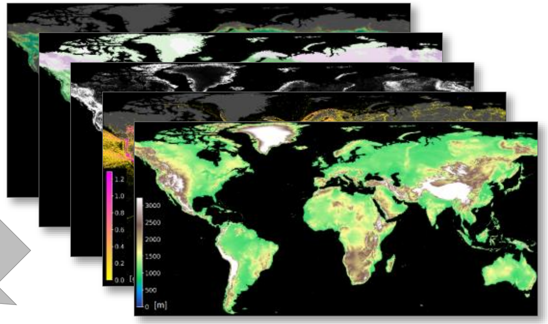
GLOBAL OZONE MAPPING

Can we apply the trained models everywhere?

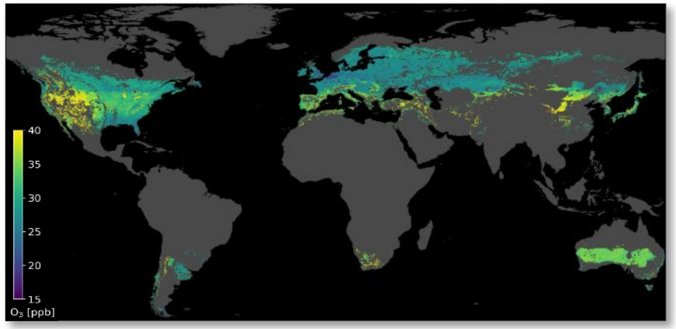


Global ozone measurements of the AQ-Bench dataset
→ Betancourt et al. 2021

Explainable Machine Learning



High-resolution geospatial data *(different sources)*



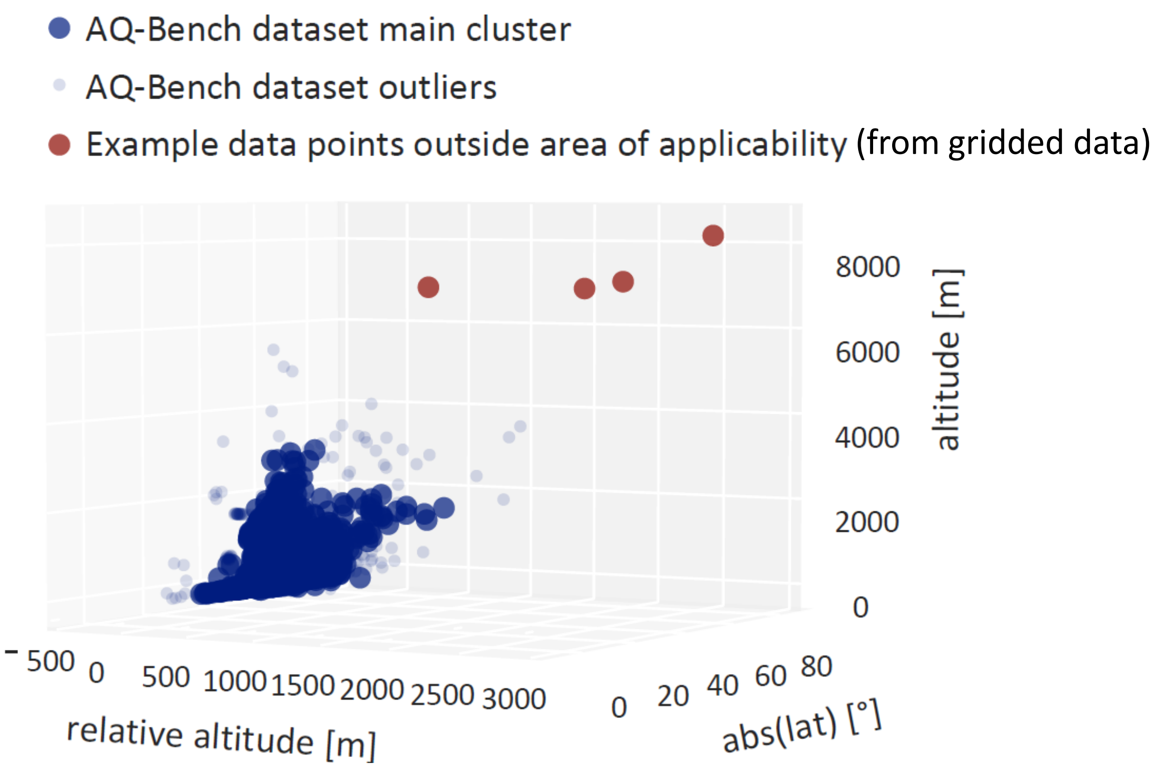
Trusted global map of ozone statistics

c.betancourt@fz-juelich.de

JÜLICH Forschungszentrum

FEATURE SPACE

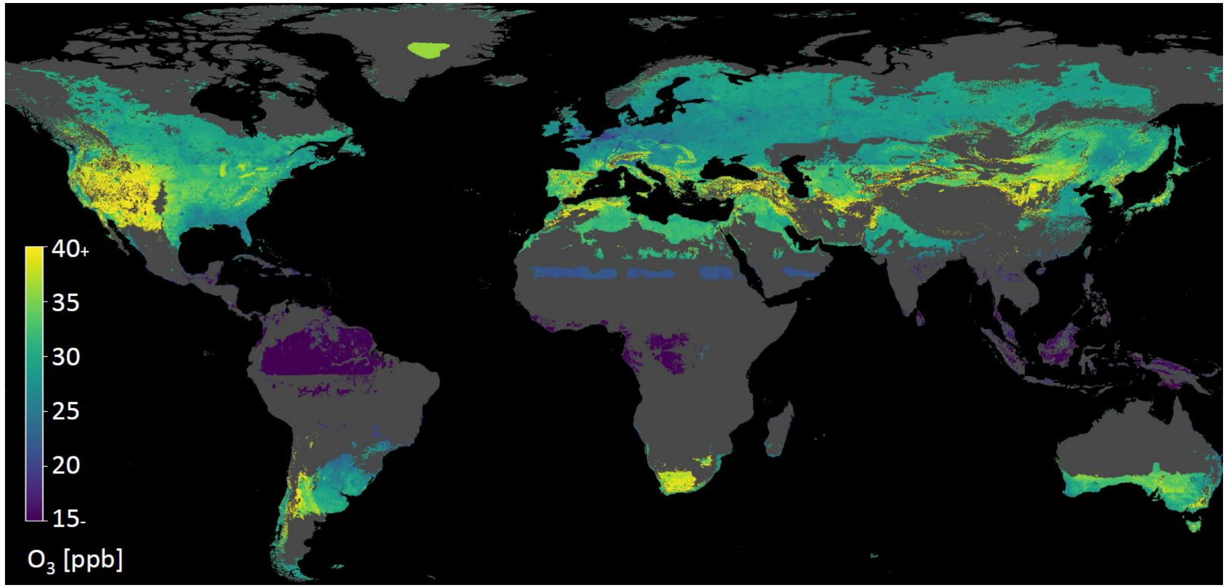
“Area of Applicability” method, Meyer et Pebesma (2021)



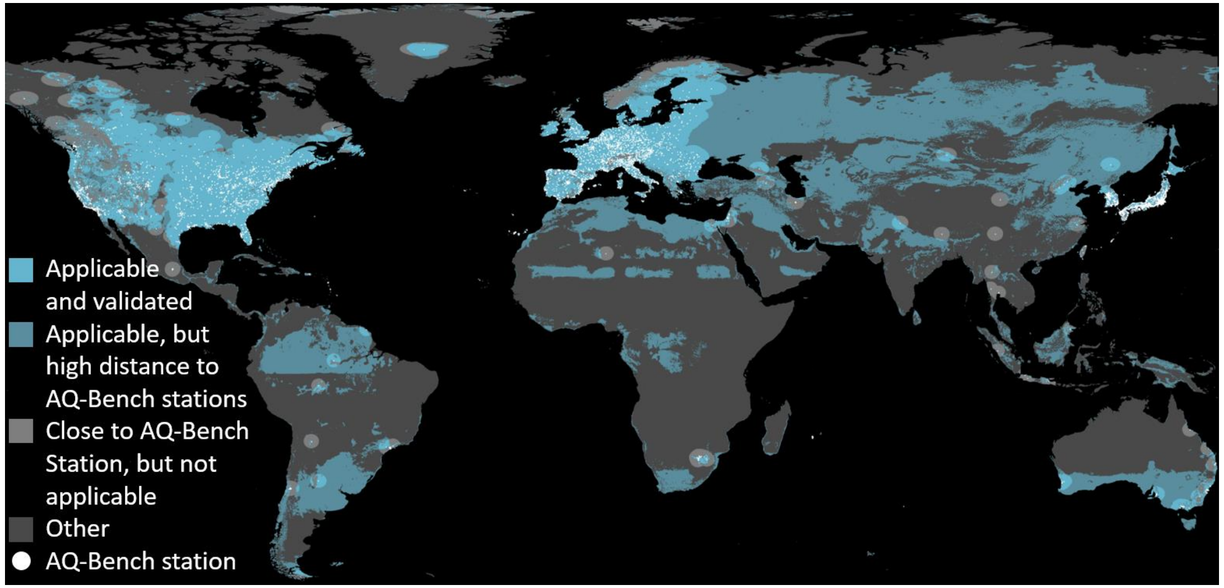
JÜLICH Forschungszentrum

GLOBAL DOMAIN

Where can we map ozone with high confidence?



Average ozone 2010-2014



Area of applicability

→ Betancourt et al. 2022

CONCLUSION

Investigating the latent space and feature space of geospatial machine learning models can lead to new insights in the global domain.

- Betancourt et al: “AQ-Bench: a benchmark dataset for machine learning on global air quality metrics” ESSD 2021, https://doi.org/10.5194/essd-13-3013-2021
- Stadtler et al: “Explainable Machine Learning Reveals Capabilities, Redundancy, and Limitations of a Geospatial Air Quality Benchmark Dataset” MAKE 2022, https://doi.org/10.3390/make4010008
- Betancourt et al: “Global, high-resolution mapping of tropospheric ozone – explainable machine learning and impact of uncertainties” GMD 2022 [preprint], https://doi.org/10.5194/gmd-2022-2

Prof. Ribana Roscher, University of Bonn

Dr. Scarlet Stadtler, Forschungszentrum Jülich

Prof. Julia Kowalski, RWTH Aachen

PD Dr. Martin Schultz, Forschungszentrum Jülich

Clara Betancourt, Forschungszentrum Jülich

Timo Tjaden Stomberg, University of Bonn

Ankit Patnala, Forschungszentrum Jülich

Ann-Kathrin Edrich, RWTH Aachen

Thank You!

JÜLICH Forschungszentrum