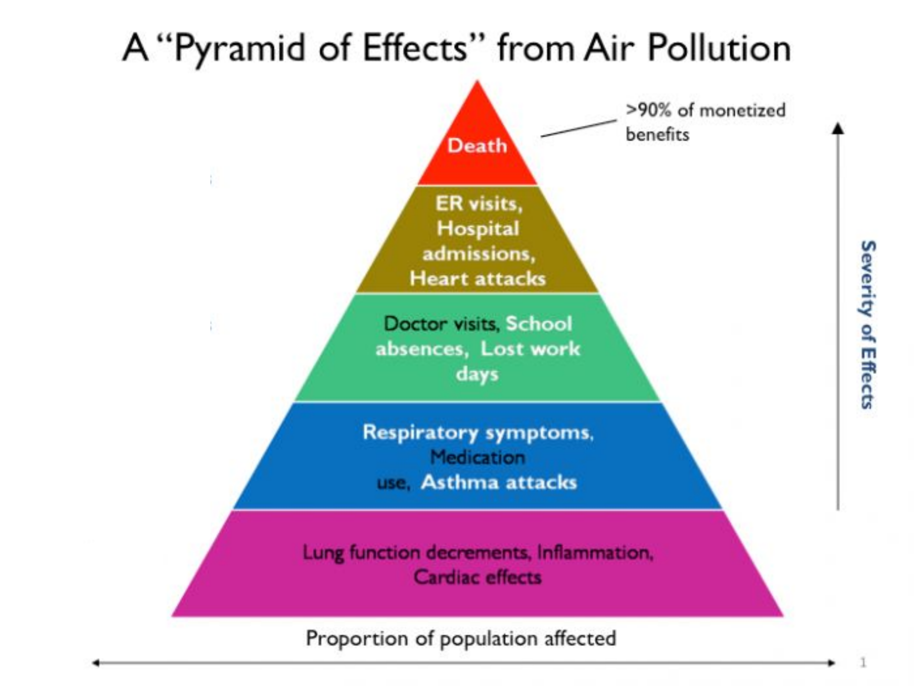
Forecasting European Ozone Air Pollution With Transformers

Seb Hickman, Paul Griffiths, Peer Nowack, Elie Alhajjar and Alex Archibald

Ozone air pollution is detrimental to human and plant health



- Between **300,000 and 1 million** estimated premature deaths worldwide annually due to ozone pollution
- WHO estimates that more than 99% of the world's population live in areas where pollutant concentrations exceed guidelines
- Several billion dollars in crop losses
- **More accurate forecasts can inform improved air quality warnings**

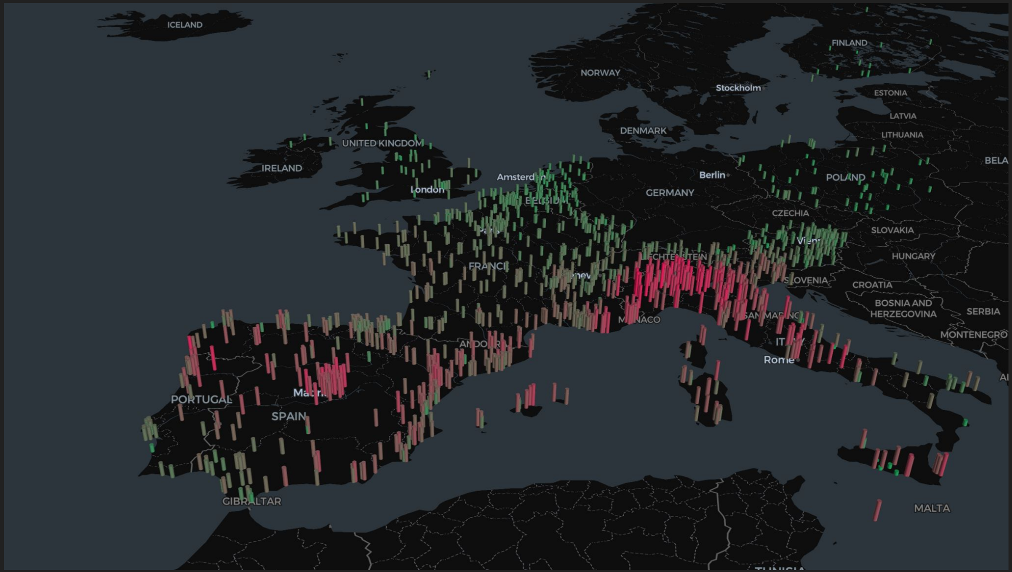
Source: EPA, https://www.epa.gov/

Why ML, and why transformers for ozone forecasting?

- Numerical methods (chemical transport models) for forecasting are computationally expensive and require parameterisations
- ML is fast and can learn complex relationships given sufficient data

- Ozone is controlled by processes which act on varying timescales
- **The temporality is important!**
- Transformers have shown state-of-the-art performance on **sequential data** in other domains such as natural language processing
- Therefore a transformer-based architecture, the Temporal Fusion Transformer, was deployed to make ozone forecasts

Surface ozone data?



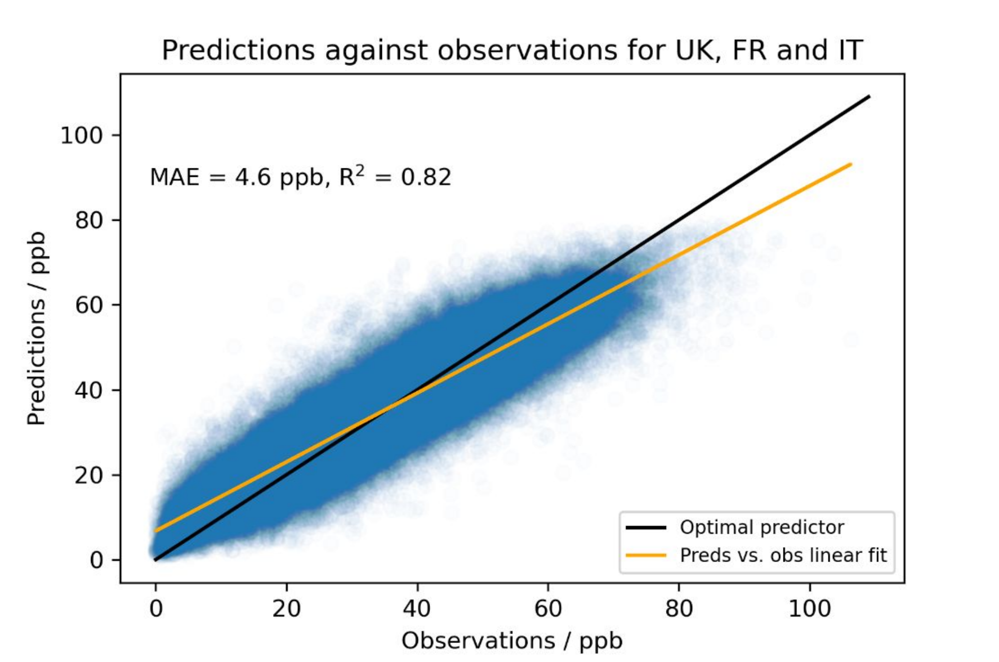
- **TOAR database! (Schultz et al.)**

- Includes **dynamic** and **static** variables at ozone measurement stations across Europe
- **Dynamic**: e.g. daily temperature (from reanalysis)
- **Static**: e.g. station population density

- Data from 1997-2014
- Train only on UK, France, and Italy
- Data split temporally, penultimate year used for validation, final year for testing

- **21 previous days of all data, and concurrent reanalysis data, used to forecast ozone 4 days ahead**

Transformer forecasts with high skill!



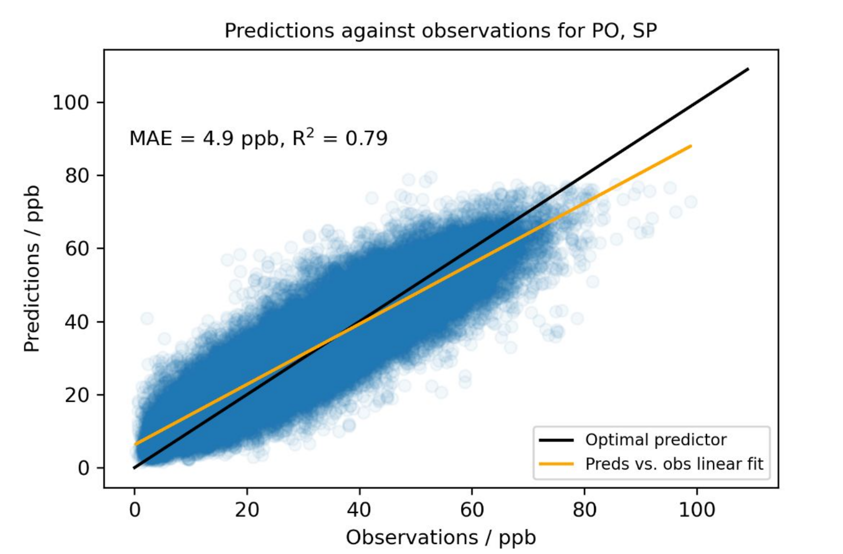
Method (and paper)	r (Pearson)	RMSE / ppb
Persistence	0.42	10.16
[32], Geos-CHEM	0.48	16.2
Ridge regression	0.50	9.59
[30], AQUM	0.64	20.8
Random forest	0.68	7.51
[33], DRR	0.70	6.3
[30], bias-corrected AQUM	0.76	16.4
[34], CNN	0.77	8.8
[23], CNN	0.79	12.0
[32], bias-corrected Geos-CHEM	0.84	7.5
LSTM	0.85	6.11
[24], RNN	0.86	12.5
[28], CNN-Transformer	*NA*	7.8
TFT	0.90	5.6

Methods in italics were tested on our dataset, while others used different data.

The difficulty of comparing methods tested on different datasets is shown by varying RMSE values.

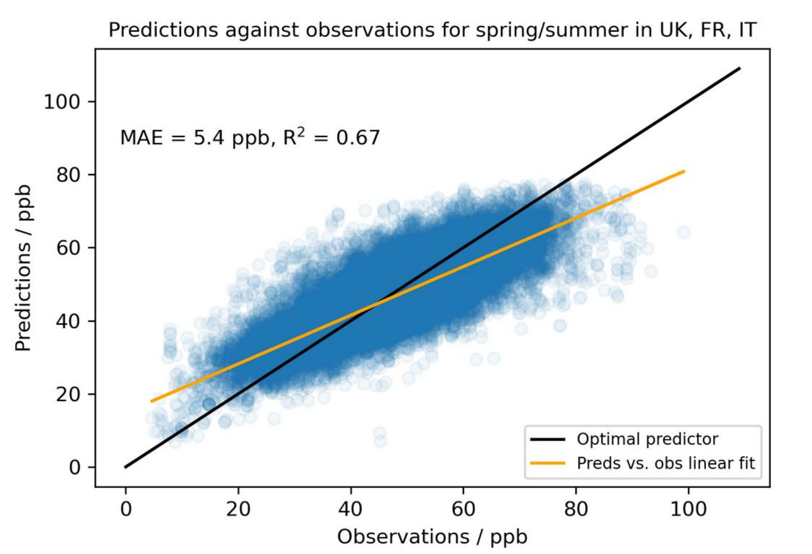
Previous 21 days of ozone and reanalysis data used to forecast 4 days ahead, with concurrent reanalysis data used for the forecast days.

Can we forecast ozone in unseen countries and at extremes?



The model generalises well to two unseen countries, Poland and Spain.

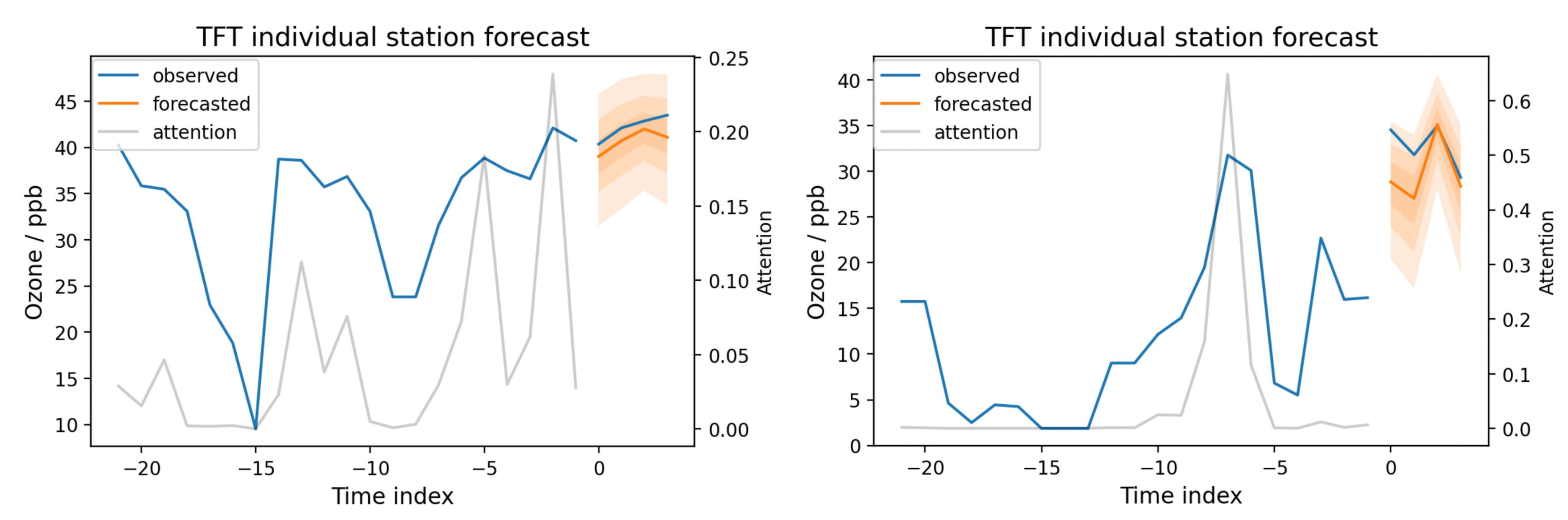
Generalisation of the transformer is better than standard ML methods (ridge regression, random forest, LSTM) on our dataset.



The model is able to make reasonable forecasts of high spring/summer ozone concentrations.

However, performance is poorer than when forecasting during the rest of the year.

How does the transformer use the previous 21 days?



- Two example forecasts at stations in our dataset
- **Attention mechanism uses previous high ozone days to help forecast future high ozone**
- Shown by the grey line illustrating the attention paid to previous days of data
- Quantile loss function used to generate prediction intervals

Overall?

- Transformer performs very skillfully in short-term ozone forecasting, outperforming standard ML methods and comparing favourably to other methods
- Performance is reasonable for high ozone concentrations, but work needed to improve performance at extremes.
- Including physical and spatial information into the model? Combination of transformers with spatial methods such as graph neural networks?
- Model interpretability is a challenge - thoughts on this are very welcome!

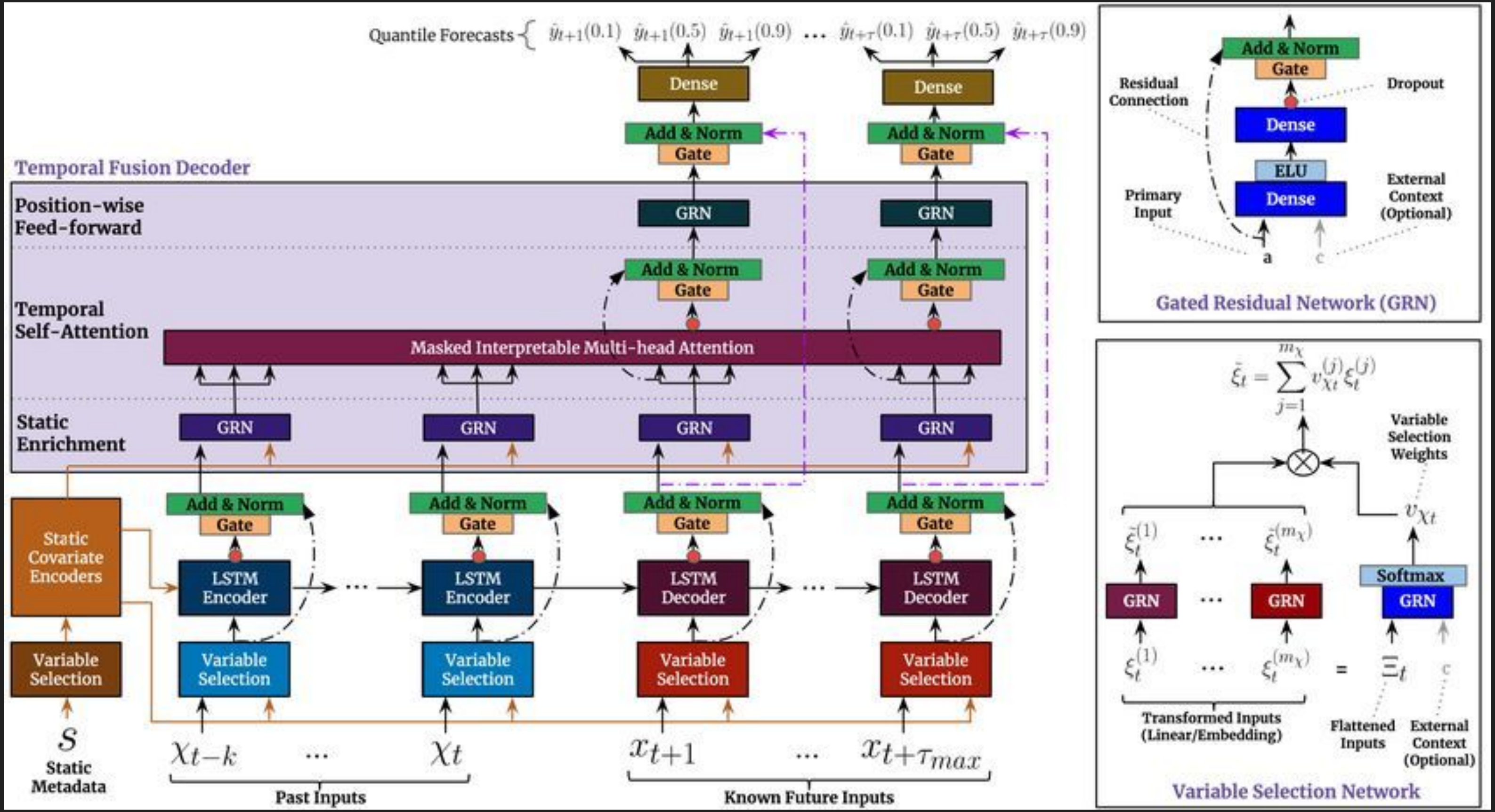
Acknowledgements

Thanks to EPSRC, AI4ER and NERC for funding this work, which was also supported by The Alan Turing Institute. The authors also thank the NERC Earth Observation Data Acquisition and Analysis Service (NEODAAS) for access to compute resources and staff support for this study.

Variables!

Variable Name	Description
Static	
station type	Characterisation of site, e.g. "background", "industrial", "traffic".
landcover	The dominant IGBP landcover classification at the station location extracted from the MODIS MCD12C1 dataset (original resolution: 0.05 degrees).
toar category	A station classification for the Tropospheric Ozone Assessment Report based on the station proxy data that are stored in the database. One of unclassified, low elevation rural, high elevation rural or urban.
pop density	Year 2010 human population per square km from CIESIN GPW v3 (original horizontal resolution: 2.5 arc minutes).
max 5km pop density	Maximum population density in a radius of 5 km around the station location.
max 25km pop density	Maximum population density in a radius of 25 km around the station location.
nightlight 1km	Year 2013 Nighttime lights brightness values from NOAA DMSP (original horizontal resolution: 0.925 km).
nightlight max 25km	Year 2013 Nighttime lights brightness values (original horizontal resolution: 5 km).
alt	Altitude of station (in m above sea level). Best estimate of the station altitude, which frequently uses the elevation from Google Earth.
station etopo alt	Terrain elevation at the station location from the 1 km resolution ETOPO1 dataset.
nox emi	Year 2010 NOx emissions from EDGAR HTAP inventory V2 in units of g m^{-2} yr^{-1} (original resolution: 0.1 degrees)
omi nox	Average 2011-2015 tropospheric NO_2 columns from OMI at 0.1 degree resolution (Env. Canada) in units of 10^{15} molecules cm^{-2}.
Dynamic	
o3	Ozone concentration, daily maximum 8-hour average statistics according to the using the EU definition of the daily 8-hour window starting from 17 h of the previous day. Measured at the station, with UV absorption.
cloudcover	Daily average cloud cover from ERA5 reanalysis for the grid cell containing a particular station.
relhum	Daily average relative humidity from ERA5 reanalysis for the grid cell containing a particular station.
press	Daily average pressure from ERA5 reanalysis for the grid cell containing a particular station.
temp	Daily average temperature from ERA5 reanalysis for the grid cell containing a particular station.
v	Daily average meridional wind speed from ERA5 reanalysis for the grid cell containing a particular station.
u	Daily average zonal wind speed from ERA5 reanalysis for the grid cell containing a particular station.
pblheight	Daily average planetary boundary layer height from ERA5 reanalysis for the grid cell containing a particular station.

Temporal Fusion Transformer (TFT)



Data ingested by the transformer?

'PAST' 'FUTURE'

Time	-21	-20	-19	-18	-17	-16	-15	-14	-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3
Ozone																						?	?	?	?
Temperature																									
Pressure																									
Wind speed																									
Station population density																									