Forecasting European Ozone Air Pollution With Transformers

Sebastian H. M. Hickman
Yusuf Hamied Department of Chemistry
University of Cambridge
Cambridge, UK, CB2 1EW
shmh4@cam.ac.uk

Paul T. Griffiths
Yusuf Hamied Department of Chemistry
University of Cambridge
Cambridge, UK, CB2 1EW
ptg21@cam.ac.uk

Peer Nowack
Climatic Research Unit
School of Environmental Sciences
University of East Anglia
Norwich, UK, NR4 7TJ
p.nowack@uea.ac.uk

Elie Alhajjar
Department of Mathematical Sciences
United States Military Academy
West Point, NY 10996
elie.alhajjar@westpoint.edu

Alex T. Archibald
Yusuf Hamied Department of Chemistry
University of Cambridge
Cambridge, UK, CB2 1EW
ata27@cam.ac.uk

Abstract

Surface ozone is an air pollutant that contributes to hundreds of thousands of premature deaths annually. Accurate short-term ozone forecasts may allow improved policy to reduce the risk to health, such as air quality warnings. However, forecasting ozone is a difficult problem, as surface ozone concentrations are controlled by a number of physical and chemical processes which act on varying timescales. Accounting for these temporal dependencies appropriately is likely to provide more accurate ozone forecasts. We therefore deploy a state-of-the-art transformer-based model, the Temporal Fusion Transformer, trained on observational station data from three European countries. In four-day test forecasts of daily maximum 8-hour ozone, the novel approach is highly skilful (MAE = 4.6 ppb, R^2 = 0.82), and generalises well to two European countries unseen during training (MAE = 4.9 ppb, R^2 = 0.79). The model outperforms standard machine learning models on our data, and compares favourably to the published performance of other deep learning architectures tested on different data. We illustrate that the model pays attention to physical variables known to control ozone concentrations, and that the attention mechanism allows the model to use relevant days of past ozone concentrations to make accurate forecasts.

1 Introduction

Surface ozone is a secondary pollutant which is not directly emitted by anthropogenic activities, but formed in the troposphere via a series of photochemical reactions [1]. Surface ozone is estimated to contribute to between 365,000 and 1.1 million premature deaths worldwide annually [2, 3, 4, 5], primarily by causing cardiovascular and respiratory diseases [6, 7, 8]. The impacts of ozone pollution

Tackling Climate Change with Machine Learning: workshop at NeurIPS 2022.

have been linked to both long and short-term exposure to high ozone [9, 10]. Background levels of ozone in remote areas often exceed guidelines, while local ozone concentrations can far exceed guidelines. The WHO estimates that 99% of the world's population live in areas where concentrations routinely exceed guidelines [11]. Due to its phytotoxicity, the negative effects of ozone air pollution on vegetation, ecosystems and crops are also significant [12, 13], leading to considerable economic losses from reduced crop yields [14]. Forecasting ozone concentrations is important to quantify and reduce the risks of ozone pollution to human and ecosystem health [15], particularly as climate change is expected to lead to increased ozone in some regions [16, 17].

Derived from the key processes controlling ozone, in-situ photochemical production and transport, there are multiple relationships between ozone and climate-sensitive environmental covariates such as temperature and meteorology that act on varying timescales [18, 19]. The contribution of each of these factors makes accurate forecasting of ozone with numerical forward (e.g. weather forecast) transport models (CTMs) and standard machine learning (ML) methods a complex and computationally expensive task. The timescale of influence of these environmental variables may be on the order of days or weeks. Varying anthropogenic emissions also affect ozone (e.g. weekdays vs. weekends), and therefore an ML approach that accounts for these temporal relationships is necessary. Transformers have been shown to be highly effective in sequential domains such as natural language processing [20, 21], in part due to their ability to attend to long-term dependencies in the data, and therefore a transformer-based model may provide an intrinsic advantage over standard ML models (such as random forests) and convolutional and recurrent neural networks that have been previously explored in the ozone forecasting literature [22, 23, 24].

In this work we train and evaluate a transformer-based ML model, focussing on the skill of the model when forecasting extreme ozone and ozone in countries unseen during training. This is a step towards ML models capable of making accurate short-term forecasts of surface ozone now, and in future climates.

2 Methodology

2.1 Model

To complement existing ML methods and numerical CTMs for ozone forecasting, we deploy a state-of-the-art temporal deep learning architecture, the Temporal Fusion Transformer (TFT) [25]. The TFT combines gated residual networks, variable selection networks, an Long-Short Term Memory (LSTM) encoder-decoder layer, and multi-head attention. The TFT ingests both static and dynamic predictive features, and in order to extract prediction intervals a quantile loss function was used.

Despite being a relatively computationally expensive ML method, training the TFT on our dataset took 2 hours using 2 Tesla V100 GPUs. Once trained, forecasts across 994 individual stations are made in seconds. This illustrates the vast speed-up of ML models compared to CTMs (which typically take hours or days to run). Hyperparameters were manually optimised for performance on validation data (see Appendix A.3).

2.2 Data

The Tropospheric Ozone Assessment Report (TOAR) dataset [26] is used as a suitable exploratory dataset for our model, due to its global coverage and large quantity of observational station ozone data. We selected data for 3 European countries: the UK, France and Italy. These were chosen to represent 3 different air quality domains, in order to test whether a single model could be trained to make accurate forecasts across domains. Data from 1997 to 2014, from all months of the year, and from 994 urban and rural stations were included in our dataset. This dataset therefore provides a larger sample of different environments than in similar previous work [22, 24, 27, 28]. Our final dataset contains more than 2 million individual days of data. We scaled our features with min-max normalisation [29].

The data include both static and dynamic features relevant to ozone concentrations. The static features relate to characteristics of a particular station, such as the local population density, while the dynamic features are environmental covariates which change through time, such as temperature. The inputs used are described in Table 2. To train, validate and test our models, the data was split temporally, with the penultimate year of station data used for validation, the final year used for testing and the

remainder for training. The previous 21 days of observations of ozone and covariates were used to make 4 day ozone forecasts. 21 previous days was chosen as an appropriate timescale for the likely temporal effects of covariates on ozone.

3 Results

When forecasting ozone concentrations using concurrently observed covariate data the model was skillful (MAE = 4.6 ppb, R^2 = 0.82, RMSE = 5.6 ppb, r = 0.90). These forecasts rely on previous ozone observations and concurrent covariate data, and therefore they are suitable for making short-term future forecasts with a meteorological forecast as input, and infilling missing ozone data in historical station data. While we cannot make direct comparisons to all published methods due to differing test datasets, the skill of our method compares favourably to other standard ML methods and numerical air quality forecasting models such as AQUM [30, 31], especially given the size and variety of our test dataset (Table 1). A correlation plot of TFT forecasts on the test set, against observations, is given in Figure 1A. The model was significantly more accurate on our data than standard ML approaches (e.g. random forests and LSTMs), and approximately 40% more accurate in MAE compared to a persistence model.

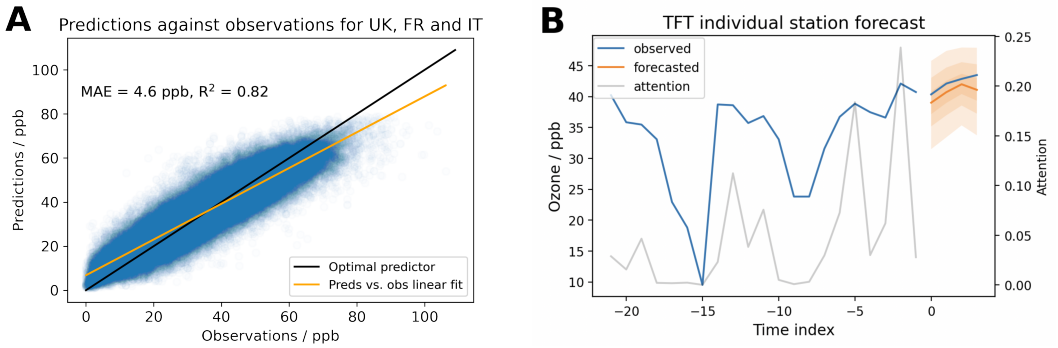

Figure 1: **A** illustrates predictions against observations on the test data for forecasting ozone with the TFT. **B** shows a 4 day forecast at a single station. The grey line shows the attention that the transformer is paying to different days in the time history. The prediction intervals generated with the quantile loss are also shown.

The skill of the TFT can be further analysed by looking at forecasts at individual stations in our dataset, as illustrated in Figure 1B. Figure 1B also shows days from the past which the attention mechanism in the model used to inform the forecast, shown by the grey line. The advantage of the transformer architecture is illustrated as the model pays attention to recent high ozone days to make future forecasts of high ozone, rather than purely the most recent days. Figure 1 also illustrates the prediction intervals generated by the TFT, which are useful to evaluate trust in the model.

The TFT's capacity to make skillful predictions of ozone concentrations at both urban and rural stations was evaluated. Comparable recent works typically focus on training and evaluating models on solely urban or rural ozone [22, 28] and it remains unclear if a single model can generalise across these 2 environments. Encouragingly, we show here that the TFT performed similarly on urban and rural data (MAE = 4.5 ppb, R^2 = 0.83 and MAE = 4.6 ppb, R^2 = 0.81, respectively). Furthermore, the feature importances of this model, derived from the attention mechanism, are largely in line with what is expected physically: both temperature and planetary boundary layer height are key variables (Appendix A.2, Figure 3).

4 Generalisation: Across Europe and Spring Ozone

To evaluate the skill of our model in generalising to unseen data, we used the model trained on data from the UK, France and Italy to make forecasts in Spain and Poland. The model was able to generalise impressively in both countries (MAE = 4.9 ppb, R^2 = 0.79), as shown in Figure 2A,

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

Table 1: The relative performance of different ML and numerical approaches when predicting ozone compared to observed ozone values. 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 the varying RMSE values.

suggesting that the model could be used to make forecasts on unseen countries without the need for extensive further training data.

To evaluate the skill of the TFT in forecasting extreme ozone concentrations, the model, trained on annual data, was evaluated on just spring and summertime ozone, when ozone concentrations in Europe tend to peak [35]. Making accurate forecasts of high ozone is important, as these high ozone concentrations pose a greater threat to health, and are likely to occur more frequently in future climates. Figure 2B also illustrates that the TFT was able to make reasonable forecasts on spring and summertime ozone concentrations (MAE = 5.4 ppb, R^2 = 0.67). However, the performance of the model evaluated on spring/summer data was poorer than performance when forecasting on data from the rest of the year.

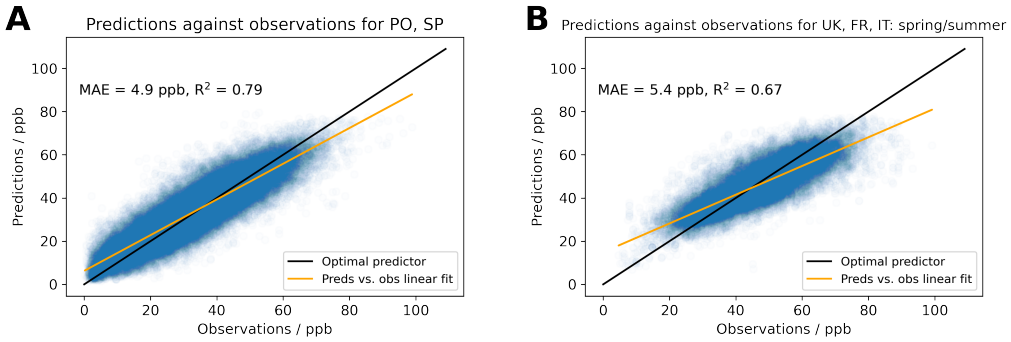

Figure 2: **A** illustrates the performance of the TFT when predicting on data from 2 countries unseen during training: Poland and Spain. **B** shows that when forecasting during spring/summertime in the UK, France and Italy, the performance of the TFT was poorer (MAE = 5.4 ppb, R^2 = 0.67) than forecasting during the rest of the year.

5 Conclusions

Forecasting ozone accurately is necessary to reduce the risk of ozone on human health now, and in future climates. However, forecasting ozone is subject to substantial numerical modelling uncertainties and is usually highly computationally expensive. As an ML alternative, a transformer-based model, the TFT, makes skillful predictions of ozone concentrations at stations across Europe. The model is able to make accurate predictions across urban and rural environments, comparing favourably to

competing methods, and performs reasonably when predicting high ozone. Promisingly, the model is able to generalise to data from 2 countries unseen in training, Poland and Spain. Our novel approach thus provides a computationally cheap method to make accurate forecasts of ozone across Europe, though further work is required to improve model predictions of extrema, such as encoding physical relationships or additional meteorological and spatial information in the model.

Acknowledgments and Disclosure of Funding

SH acknowledges funding from EPSRC via the AI4ER CDT at the University of Cambridge (EP/S022961/1), and support from The Alan Turing Institute. PTG and ATA were financially supported by NERC through NCAS (R8/H12/83/003). PN was supported through an Imperial College Research Fellowship. The authors thank the NERC Earth Observation Data Acquisition and Analysis Service (NEODAAS) for access to compute resources and staff support for this study.

References

[1] Barbara J Finlayson-Pitts and James N Pitts Jr. Tropospheric air pollution: ozone, airborne toxics, polycyclic aromatic hydrocarbons, and particles. *Science*, 276(5315):1045–1051, 1997.

[2] Christopher JL Murray, Aleksandr Y Aravkin, Peng Zheng, Cristiana Abbafati, Kaja M Abbas, Mohsen Abbasi-Kangevari, Foad Abd-Allah, Ahmed Abdelalim, Mohammad Abdollahi, Ibrahim Abdollahpour, et al. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. *The Lancet*, 396(10258):1223–1249, 2020.

[3] Susan C Anenberg, Larry W Horowitz, Daniel Q Tong, and J Jason West. An estimate of the global burden of anthropogenic ozone and fine particulate matter on premature human mortality using atmospheric modeling. *Environmental Health Perspectives*, 118(9):1189–1195, 2010.

[4] Christopher S Malley, Daven K Henze, Johan CI Kuylenstierna, Harry W Vallack, Yanko Davila, Susan C Anenberg, Michelle C Turner, and Mike R Ashmore. Updated global estimates of respiratory mortality in adults 30 years of age attributable to long-term ozone exposure. *Environmental Health Perspectives*, 125(8):087021, 2017.

[5] Raquel A Silva, J Jason West, Yuqiang Zhang, Susan C Anenberg, Jean-François Lamarque, Drew T Shindell, William J Collins, Stig Dalsoren, Greg Faluvegi, Gerd Folberth, et al. Global premature mortality due to anthropogenic outdoor air pollution and the contribution of past climate change. *Environmental Research Letters*, 8(3):034005, 2013.

[6] Sun-Young Kim, Esther Kim, and Woo Jin Kim. Health effects of ozone on respiratory diseases. *Tuberculosis and Respiratory Diseases*, 83(Supple 1):S6, 2020.

[7] EC Filippidou and A Koukouliata. Ozone effects on the respiratory system. *Prog Health Sci*, 1(2), 2011.

[8] Haitong Zhe Sun, Pei Yu, Changxin Lan, Michelle WL Wan, Sebastian Hickman, Jayaprakash Murulitharan, Huizhong Shen, Le Yuan, Yuming Guo, and Alexander T Archibald. Cohort-based long-term ozone exposure-associated mortality risks with adjusted metrics: A systematic review and meta-analysis. *The Innovation*, page 100246, 2022.

[9] Michelle L Bell, Aidan McDermott, Scott L Zeger, Jonathan M Samet, and Francesca Dominici. Ozone and short-term mortality in 95 us urban communities, 1987-2000. *Jama*, 292(19):2372–2378, 2004.

[10] Daniela Nuvolone, Davide Petri, and Fabio Voller. The effects of ozone on human health. *Environmental Science and Pollution Research*, 25(9):8074–8088, 2018.

[11] WHO and ECE. *WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. World Health Organization, 2021.

[12] David Fowler, Kim Pilegaard, MA Sutton, Per Ambus, M Raivonen, Jan Duyzer, David Simpson, H Fagerli, Sandro Fuzzi, JK Schjoerring, et al. Atmospheric composition change: ecosystems–atmosphere interactions. *Atmospheric Environment*, 43(33):5193–5267, 2009.

[13] LD Emberson, MR Ashmore, D Simpson, J-P Tuovinen, and HM Cambridge. Modelling and mapping ozone deposition in europe. *Water, Air, and Soil Pollution*, 130(1):577–582, 2001.

[14] Jennifer Burney and V Ramanathan. Recent climate and air pollution impacts on indian agriculture. *Proceedings of the National Academy of Sciences*, 111(46):16319–16324, 2014.

[15] JL Schnell, CD Holmes, A Jangam, and MJ Prather. Skill in forecasting extreme ozone pollution episodes with a global atmospheric chemistry model. *Atmospheric Chemistry and Physics*, 14(15):7721–7739, 2014.

[16] Hang Lei, Donald J Wuebbles, and Xin-Zhong Liang. Projected risk of high ozone episodes in 2050. *Atmospheric Environment*, 59:567–577, 2012.

[17] Flossie Brown, Gerd A Folberth, Stephen Sitch, Susanne Bauer, Marijn Bauters, Pascal Boeckx, Alexander W Cheesman, Makoto Deushi, Inês Dos Santos Vieira, Corinne Galy-Lacaux, et al. The ozone–climate penalty over south america and africa by 2100. *Atmospheric Chemistry and Physics*, 22(18):12331–12352, 2022.

[18] Ibai Laña, Javier Del Ser, Ales Padró, Manuel Vélez, and Carlos Casanova-Mateo. The role of local urban traffic and meteorological conditions in air pollution: A data-based case study in madrid, spain. *Atmospheric Environment*, 145:424–438, 2016.

[19] Xiang Weng, Grant Forster, and Peer Nowack. A machine learning approach to quantify meteorological drivers of recent ozone pollution in china. *Atmospheric Chemistry and Physics Discussions*, pages 1–28, 2022.

[20] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901, 2020.

[21] Yanrong Ji, Zhihan Zhou, Han Liu, and Ramana V Davuluri. Dnabert: pre-trained bidirectional encoder representations from transformers model for dna-language in genome. *Bioinformatics*, 37(15):2112–2120, 2021.

[22] Felix Kleinert, Lukas H Leufen, and Martin G Schultz. Intellio3-ts v1. 0: a neural network approach to predict near-surface ozone concentrations in germany. *Geoscientific Model Development*, 14(1):1–25, 2021.

[23] Ebrahim Eslami, Yunsoo Choi, Yannic Lops, and Alqamah Sayeed. A real-time hourly ozone prediction system using deep convolutional neural network. *Neural Computing and Applications*, 32(13):8783–8797, 2020.

[24] Fabio Biancofiore, Marco Verdecchia, Piero Di Carlo, Barbara Tomassetti, Eleonora Aruffo, Marcella Busilacchio, Sebastiano Bianco, Sinibaldo Di Tommaso, and Carlo Colangeli. Analysis of surface ozone using a recurrent neural network. *Science of the Total Environment*, 514:379–387, 2015.

[25] Bryan Lim, Sercan Ö Arık, Nicolas Loeff, and Tomas Pfister. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4):1748–1764, 2021.

[26] Martin G Schultz, Sabine Schröder, Olga Lyapina, Owen R Cooper, Ian Galbally, Irina Petropavlovskikh, Erika Von Schneidemesser, Hiroshi Tanimoto, Yasin Elshorbany, Manish Naja, et al. Tropospheric ozone assessment report: Database and metrics data of global surface ozone observations. *Elementa: Science of the Anthropocene*, 5, 2017.

[27] Jun Ma, Zheng Li, Jack CP Cheng, Yuexiong Ding, Changqing Lin, and Zherui Xu. Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network. *Science of The Total Environment*, 705:135771, 2020.

[28] Yibin Chen, Xiaomin Chen, Ailan Xu, Qiang Sun, and Xiaoyan Peng. A hybrid cnn-transformer model for ozone concentration prediction. *Air Quality, Atmosphere & Health*, pages 1–14, 2022.

[29] T Jayalakshmi and A Santhakumaran. Statistical normalization and back propagation for classification. *International Journal of Computer Theory and Engineering*, 3(1):1793–8201, 2011.

[30] LS Neal, P Agnew, S Moseley, C Ordóñez, NH Savage, and M Tilbee. Application of a statistical post-processing technique to a gridded, operational, air quality forecast. *Atmospheric Environment*, 98:385–393, 2014.

[31] Ulas Im, Roberto Bianconi, Efisio Solazzo, Ioannis Kioutsioukis, Alba Badia, Alessandra Balzarini, Rocío Baró, Roberto Bellasio, Dominik Brunner, Charles Chemel, et al. Evaluation of operational on-line-coupled regional air quality models over europe and north america in the context of aqmeii phase 2. part i: Ozone. *Atmospheric Environment*, 115:404–420, 2015.

[32] Peter D Ivatt and Mathew J Evans. Improving the prediction of an atmospheric chemistry transport model using gradient-boosted regression trees. *Atmospheric Chemistry and Physics*, 20(13):8063–8082, 2020.

[33] Edouard Debry and Vivien Mallet. Ensemble forecasting with machine learning algorithms for ozone, nitrogen dioxide and pm10 on the prev'air platform. *Atmospheric Environment*, 91:71–84, 2014.

[34] Alqamah Sayeed, Yunsoo Choi, Ebrahim Eslami, Yannic Lops, Anirban Roy, and Jia Jung. Using a deep convolutional neural network to predict 2017 ozone concentrations, 24 hours in advance. *Neural Networks*, 121:396–408, 2020.

[35] Alastair Lewis, James Allan, David Carruthers, David Carslaw, Gary Fuller, Roy Harrison, Mat Heal, Eiko Nemitz, and Claire Reeves. Ozone in the uk – recent trends and future projections. *Department for Environment, Food and Rural Affairs Report*, 2021.

Data and code availability

The TOAR dataset is publicly available online [26], and code used in this work will be made publicly available.

A Appendices

A.1 Features from the TOAR dataset

Table 2 describes the data used as features for the machine learning model. The features are split into static and dynamic features. Static features describe the characteristics of a particular station, while dynamic features vary through time. Due to the large size and relative completeness of our dataset, imputing missing values was deemed unnecessary, and rows with missing data were dropped.

A.2 Feature importances

The feature importances of the TFT, derived from the weights of attention mechanism in our model, are shown in Figure 3. These importances are largely in line with what is expected physically: both temperature and planetary boundary layer height are key variables.

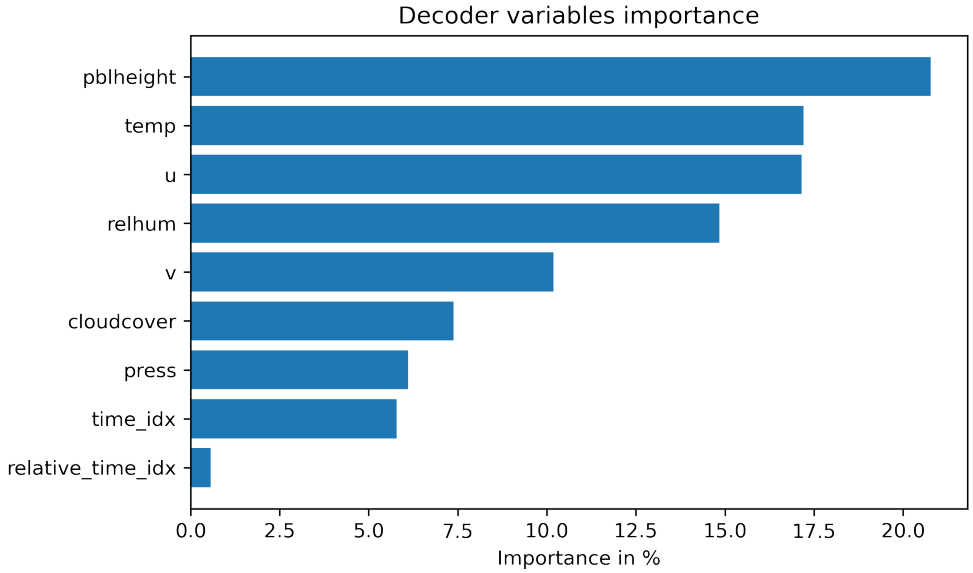

Figure 3: The variable importances of the TFT when making forecasts, derived from the weights of the attention mechanism. These are largely in line with expected physical relationships.

A.3 Model hyperparameters

Table 3 details the hyperparameters used for the TFT model. These hyperparameters were selected with manual optimisation, however more principled methods such a random search or Bayesian optimisation will be implemented in future work.

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.

Table 2: Table giving the relevant data extracted from the TOAR database.

Model	Hyperparameter value
TFT	
attention head size	4
dropout	0.2
hidden continuous size	16
hidden size	32
learning rate	0.0302
lstm layers	2
optimizer	ranger

Table 3: Table giving the hyperparameters for the final TFT used for model evaluation, determined by manual optimisation. During manual optimisation, larger hidden and attention head sizes were tested, however increasing these values did not improve performance.