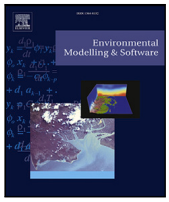
Environmental Modelling and Software 177 (2024) 106048

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft



Transfer learning in environmental data-driven models: A study of ozone forecast in the Alpine region

Matteo Sangiorgio *, Giorgio Guariso

Department of Electronics, Information and Bioengineering, Politecnico di Milano, Via Ponzio 34/5, I-20133, Milan, Italy

ARTICLE INFO

Keywords:
Air pollution forecasting
Nonlinear dynamics
Machine learning
LSTM neural networks
Multi-step prediction

ABSTRACT

Many environmental variables, in particular, related to air or water quality, are measured in a limited number of points and often for a limited time span. This forbids the development of accurate models for interesting locations with missing or insufficient data and poses the question of whether a model developed for another measurement site can be reliably applied. Such a question is particularly critical when the model is entirely data-driven, such as a neural network. In this context, the paper proposes a procedure to evaluate the expected performance of an existing neural network model applied to a new unmonitored station. This transferability assessment is exemplified by the problem of forecasting ozone concentrations in different environmental settings around the Alpine Arc. Long Short-Term Memory (LSTM) neural network models are applied for predicting hourly concentrations in 20 stations of different types (urban, rural, and mountain). The analysis of the results allows us to determine the expected performance of such models in new cases and reduce the transferability uncertainty when the existing models can be partitioned into clusters. The LSTM models demonstrate the possibility of high accuracy in ozone forecasting at all sites. Given the significant impacts of this gas on human health and the environment, this can contribute to better decision-making and mitigation strategies for air pollution control.

Software and data availability

The LSTM neural network for the ozone prediction was implemented in python (version 3.8) using PyTorch library. Authors' experimental environment was as follows:

- OS: Ubuntu 20.04.5 LTS
- CPU: Intel(R) Core(TM) i9-10900K CPU @ 3.70 GHz (20 threads)
- RAM: 32 GB
- GPU: NVIDIA Quadro RTX 5000 (16 GB)

Alpine ozone data were collected by ARPA Lombardia, ARPA Piemonte, ARPA Valle d'Aosta, Agenzia provinciale per l'ambiente e la tutela del clima of Bozen, Swiss Federal Office for the Environment, ATMO Auvergne-Rhône-Alpes, and Amt der Tiroler Landesregierung.

1. Introduction

The application of a model developed for a specific site to other situations where the necessary local data are missing or insufficient has been studied in environmental sciences for at least half a century (see, for instance, Cordery, 1971; O'Donnell, 1976; Gauch et al., 2021). Measurements essential to support decision-making are often limited to a few sites, whereas more detailed knowledge would be necessary. One must thus apply models developed elsewhere when available data are insufficient. The problem is best defined for physically-based models, which rely on theoretical knowledge rather than solely on the data. In this case, the number of parameters is often limited, and, more importantly, they have a precise physical meaning, or they may even be directly measurable on the field. Conversely, if the model is fully data-based, as for neural networks, it is challenging to devise which specific domain features allow extending a model developed for another similar case. The problem is known as "transfer learning" in the neural network literature and has been extensively studied in recent years (see, for instance, Pan and Yang, 2009; Yosinski et al., 2014; Long et al., 2015; Glynis et al., 2023; Kerimov et al., 2023). Machine learning traditionally assumes that identification and application datasets are different realizations of the same stochastic process and, thus, have the same statistical distribution. Recent research, however, has shown that machine learning models, such as neural networks, can have a high generalization capability, *i.e.*, a wide transferability (Glorot et al., 2011), and strongly limit the performance degradation caused by the

* Corresponding author.
E-mail address: matteo.sangiorgio@polimi.it (M. Sangiorgio).

https://doi.org/10.1016/j.envsoft.2024.106048
Received 17 December 2023; Received in revised form 12 April 2024; Accepted 17 April 2024
Available online 25 April 2024
1364-8152/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

discrepancy across domain distributions in many applications (Yosinski et al., 2014). This confirms the possibility of storing knowledge acquired by solving one task and applying it to another similar problem.

Within this field, the scientific literature has mainly considered the problem of "domain adaptation". It determines how to modify the model developed for a specific source domain to apply it to a new target domain. While there are different techniques to accomplish the adaptation (e.g., Glorot et al., 2011; You et al., 2019; Farahani et al., 2021; Yao et al., 2022), they are invariably based on the availability of some information on the target domain, which allows for mapping its characteristics to those of the source domain (see, for instance, Ghosh et al., 2022; Peng et al., 2022; Willard et al., 2021). Applications of these methods have also been proposed in the case of environmental data-driven models (Guariso et al., 2020; Himeur et al., 2022; Yuan et al., 2022; Xu et al., 2023; Priyatikanto et al., 2023; Yang et al., 2023; Kerimov et al., 2023), in particular for air pollution (Sangiorgio and Guariso, 2023; Ma et al., 2022; Fong et al., 2020; Ma et al., 2019).

This study tackles the problem from a different perspective and does not assume any information on the new target domain (i.e., the new measurement site) is available. We assess, in fact, the expected performances of an existing model to a new case, that is, the transferability of the model, by analysing the model's performance in other similar cases. The approach is tested on the case of multi-step neural forecasting of ozone concentration, characterized by a complex dynamic which has proved to be chaotic (Chattopadhyay and Chattopadhyay, 2008). It thus represents a challenging testing ground, widely used in the past to assess a range of different applications, such as prediction (Abdul-Wahab and Al-Alawi, 2002; Lee et al., 2004; Sfetsos and Siriopoulos, 2004; Pisoni et al., 2009; Biancofiore et al., 2015; Maciąg et al., 2019), imputation of missing data (Camastra et al., 2022; Betancourt et al., 2023), and variable selection (Diao et al., 2013; Alijla et al., 2018; Zhang et al., 2022).

Air pollution, notably ozone pollution, is a persistent environmental issue affecting many areas worldwide (Zhang et al., 2019; Tai et al., 2014). The accurate prediction of ozone concentrations is crucial for effective air quality management and the implementation of appropriate mitigation measures. Traditional prediction methods often rely on statistical models, which may have limitations in capturing the complex temporal dependencies and nonlinear relationships in ozone concentration data. Deep learning techniques, such as LSTM neural networks, have shown promise in time series forecasting tasks in recent years. This paper thus applies Long Short-Term Memory (LSTM) neural network models for predicting hourly ozone concentrations in twenty urban, rural, and mountain stations around the Alpine Arc. The purpose of the paper is not, however, the discussion of forecasting models' architectures, for which a vast literature already exists (see, for instance, the surveys in Méndez et al., 2023, and Masood and Ahmad, 2021), but to evaluate the feasibility of transferring the developed models across different sites.

The remainder of the paper is organized as follows. Section 2 formalizes the idea of transferability of data-driven models and describes the proposed framework to derive transferability indicators. In Section 3, we introduce the Alpine ozone case study, analyse the available dataset, and define the predictive task. In Section 4, a comprehensive overview of the results is given, and the key insights are discussed. Some concluding remarks are drawn in Section 5.

2. Assessing transferability

As anticipated in the introduction, we consider the most restrictive transfer learning approach, which consists of identifying a machine learning model for a data-rich site and assessing its transferability without retraining or tuning to another data-scarce site. In other words, suppose a set of models can be trained, validated and tested for a limited number of the source domains (sites si, $i = 1, \dots, N$ with $N > 1$), where sufficient observed data are available. We want to assess the accuracy of one such model when used for a new site sh. Assuming such accuracy is measured by a suitable performance indicator P, the problem is how to determine the value $P(si, sh)$ (i.e., the performance on site sh of a model trained on site si) based on the computed values $P(si, sj)$, $j = 1, \dots, N$ (see Fig. 1).

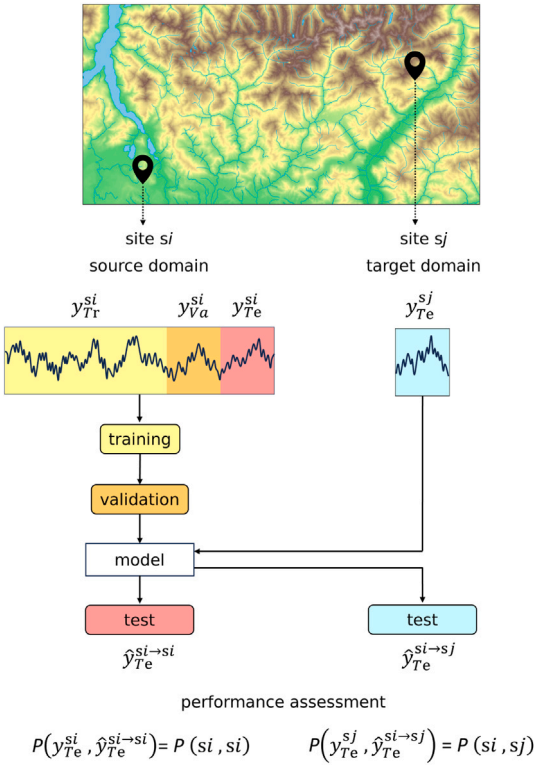


Fig. 1. Schematic representation of spatial transfer learning: si is the source domain (used for training, validation and test), while sj is the target domain (only test). $P(\cdot,\cdot)$ is the performance indicator that quantifies the similarity between actual and predicted data.

When only a single site is considered, the performance assessment is that of a traditional machine learning task, i.e., without transfer learning ($P(si, si)$ in the current notation).

The transferability problem can be solved with the following procedure (see Fig. 2).

- Specify the modelling task (forecasting, simulation, etc.) and develop a model (i.e., train, validate and test) for that task for the available sites.
- Cross-test the performance of each model developed for a specific site (that acts as a source domain) to the other sites (target domains) without retraining and only computing the desired performance index. A transferability matrix is generated, considering all the possible combinations.
- Compute two (global) transferability indicators: the average performance of all the sites on themselves, $P(si, si)$ (elements on the diagonal of the transferability matrix) and the average of $P(si, sj)$, with $i \neq j$ (off-diagonal elements).
- In the case of several source domains with heterogeneous behaviours, a clustering algorithm can be applied to the transferability matrix to group the sites into a (limited) number of clusters with different characteristics. Each cluster is associated with an intra-cluster transferability sub-matrix. The algorithm aggregates the sites so that the variability of the off-diagonal elements of each sub-matrix is minimized (this can be done, in simple cases, by considering all the possible combinations, or, more in general, adopting a combinatorial optimization algorithm).

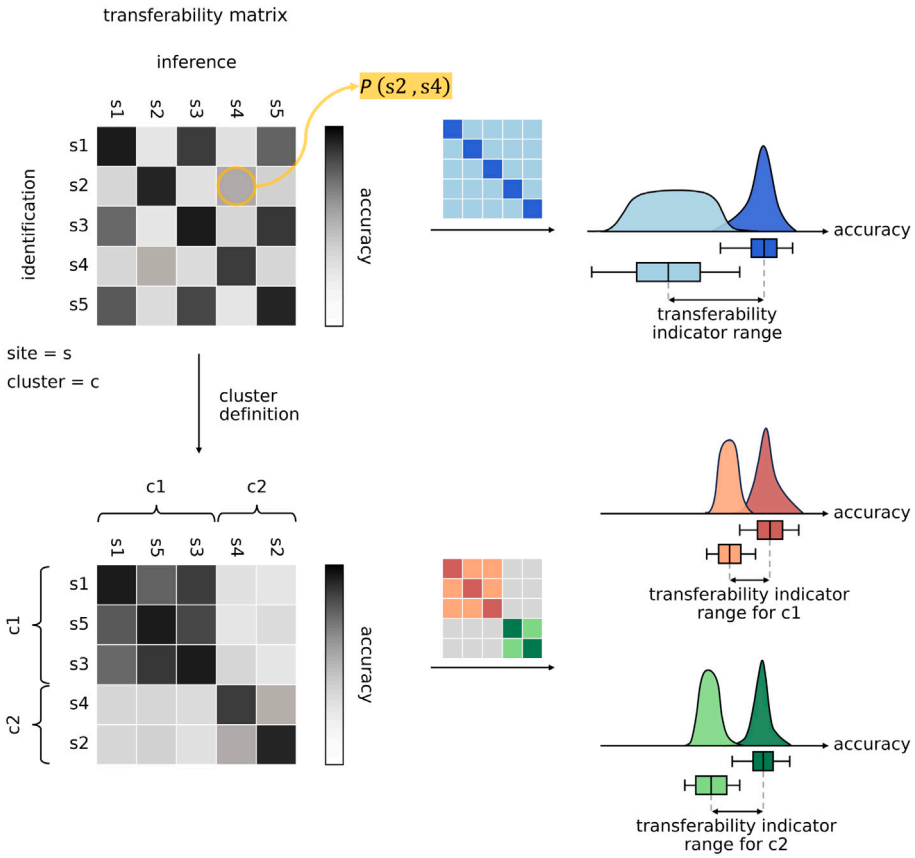


Fig. 2. Outline of the transferability assessment framework. The transferability matrix is computed by cross-testing the performance of a model developed for a specific site to the other sites. From the matrix, a (global) transferability indicator range can be obtained. If the transferability matrix shows a tendency to cluster, a more precise transferability indicator range can be computed for each group. The boxes indicate the 25-percentile, average and 75-percentile; whiskers cover the whole distribution range.

- Compute the intra-cluster transferability indicators by averaging the transferability indicators for each sub-matrix.

The procedure's output is one (or more) range(s) of values, i.e., the transferability indicator range, that allows a quantification of the reliability of the model developed for a site to other new sites for which a sufficient dataset is not (yet) available. When several models are available, the complete distribution of the self-performance and cross-performance can be computed (see the distributions and the corresponding box and whiskers plots in Fig. 2). One can thus select any quantile of the distribution instead of the average value. If a new site can be assigned (for some external features) to the correct cluster, the reliability is indicated by the intra-cluster range. In the opposite case, in which the new station is incorrectly assigned to another cluster, the global indicator quantifies the accuracy one can expect by applying any of the developed models.

3. Alpine ozone case study

The procedure presented above has been applied to the problem of forecasting hourly ozone concentrations in twenty different sites around the Alpine Arc. They include measurement stations from four European countries (Austria, France, Italy, and Switzerland) as shown in the map of Fig. 3.

At ground level, ozone (O_3) is a well-known pollutant that has detrimental effects on human health and vegetation. High concentrations can damage the respiratory system and exacerbate cardiovascular, reproductive, metabolic, and central nervous system problems (Paoletti et al., 2007; Stowell et al., 2017; US EPA, 2020). The concentration level recommended by the World Health Organization (WHO) is below 100 μg m^{-3} (in terms of mean values of the daily maximum 8-hour average, MDA8) (Wang et al., 2017), while other targets on yearly indicators such as AOT40 or SOMO35 are set by the European Union for the protection of vegetation and human health (Guerreiro et al., 2014; Cakaj et al., 2023).

O_3 is a secondary pollutant, i.e., its concentrations derive from the emissions of "precursors" gases. In particular, ozone derives from complex chemical and physical highly nonlinear reactions mainly involving nitrogen oxides and volatile organic compounds (VOC) activated by ultraviolet radiation (Lin et al., 1988; Chen et al., 1998; Zolghadri et al., 2004; Agirre-Basurko et al., 2006; Singh et al., 2011). This means that a non-negligible presence of ozone is usually possible during daylight and mainly in summer, thus following a cyclical variation over periods of 24 h and 365 days. The main sources of nitrogen oxides (NOx) are usually road transport and domestic heating, while VOCs are due to industrial plants, road transport, and agriculture (Finlayson-Pitts and Pitts, 1993; Wang et al., 2018). The ozone formation process may take a few hours, and high concentrations may be found at kilometres of distance from the precursors' sources, depending on the meteorological conditions (Lelieveld and Dentener, 2000). In particular, higher concentrations may develop at high elevations because of the higher sun radiation.

3.1. Data analysis

Data for the experiment refer to 20 stations in the Alpine Arc and were collected from different sources, all referring to the period 2007–2021 with an hourly time step. For the Italian side, values

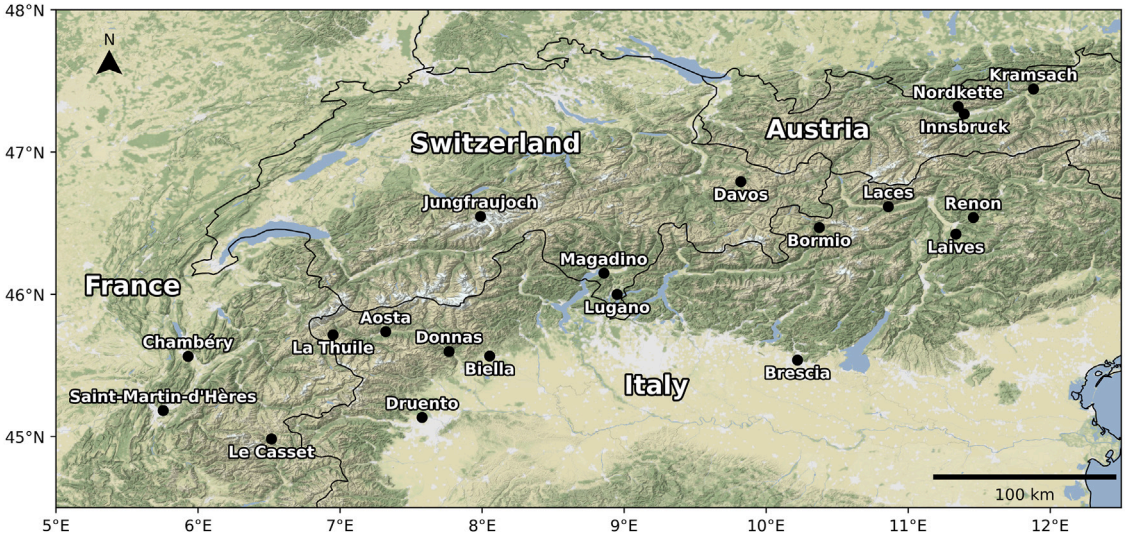


Fig. 3. Position of the 20 ozone concentration monitoring stations.

Table 1
Features of the considered ozone monitoring stations.

Category	Station	Country	Elevation [m asl]	Population	Population Density [per km^2]	Missing data [%]
Urban	Aosta	Italy	583	33 093	1 548.6	1.14
	Biella	Italy	420	44 324	949.5	4.00
	Brescia	Italy	149	200 423	2 219.5	4.78
	Laives	Italy	258	17 168	706.5	2.61
	Chambéry	France	270	59 172	2 819.1	4.10
	Saint-Martin-d'Hères	France	234	38 188	4 124.0	2.72
	Innsbruck	Austria	574	132 493	1 262.9	4.26
	Lugano	Switzerland	273	63 185	832.1	0.36
Rural	Donnas	Italy	322	2 420	71.2	2.97
	Druento	Italy	285	8 863	321.8	6.58
	Laces	Italy	639	5 156	65.4	2.98
	Kramsach	Austria	602	4 891	181.8	4.36
	Magadino	Switzerland	196	1 604	219.7	0.39
Mountain	La Thuile	Italy	1 450	790	6.3	2.63
	Bormio	Italy	1 225	4 088	99.7	3.50
	Renon	Italy	1 154	7 925	71.4	2.25
	Le Casset (Le Monêtier-les-Bains)	France	1 526	1 062	10.9	3.62
	Nordkette (Innsbruck)	Austria	1 958	–	–	4.33
	Davos	Switzerland	1 560	10 898	38.4	0.41
	Jungfraujoch	Switzerland	3 463	–	–	1.66

come from the routine measurements by the regional Environmental Agencies: Brescia and Bormio are from Lombardy (ARPA Lombardia), Biella and Druento from Piedmont (ARPA Piemonte), Aosta and Donnas from Aosta Valley (ARPA Valle d'Aosta), and Laces, Laives and Renon from South Tyrol (Agenzia provinciale per l'ambiente e la tutela del clima of Bozen). The data of the Swiss stations, Lugano, Magadino, Davos and Jungfraujoch, are stored in the NABEL database of the Federal Office for the Environment (FOEN). Concentrations of the French stations, Chambèry, Saint-Martin-d'Hères, and Le Casset, were provided by ATMO Auvergne-Rhône-Alpes and those of the Austrian stations, Innsbruck, Kramsach, and Nordkette, by the Amt der Tiroler Landesregierung. In all cases, the data missing in the time series of over 131,000 values for each station are very limited (from 0,39 to 6,58% with an average of 3%). Table 1 reports the most significant descriptors of the measurement sites. Each dataset was preprocessed by removing missing values and splitting it into training (2007–2015), validation (2016–2018), and test (2019–2021) sets.

As shown in Table 1, the stations are traditionally subdivided into those representing urban environments (red in the following figures), within towns of tens of thousands of inhabitants, rural (green), in areas with a population density of the order few hundred inhabitants per square kilometre, and mountain sites (blue), at elevations above 1000 m with a rather sparse population. An exception to this classification is represented by Bormio (IT) and Davos(CH), well-known holiday sites that can double or triple their population because of tourists in summer and winter. In this case, the resident population in Table 1 does not fully reflect the level of urbanization and traffic.

Fig. 4 shows the distributions of the concentration values of all stations. It appears that the medians (white circles) are generally higher in mountain sites coherently with the higher solar radiation. Another relevant feature is that the variability (see the black bars indicating the range of 25%–75% of the values) is much higher in urban and rural sites than in the mountains. Indeed, all urban and rural sites have a large portion of the values close to zero, while in the mountain sites, concentrations are almost always greater than 50 $\mu g\ m^{-3}$. Bormio is slightly different for the reasons already explained.

Figs. 5 and 6 show other characteristic features of the data set. Mountain sites are much less influenced by the diurnal and annual

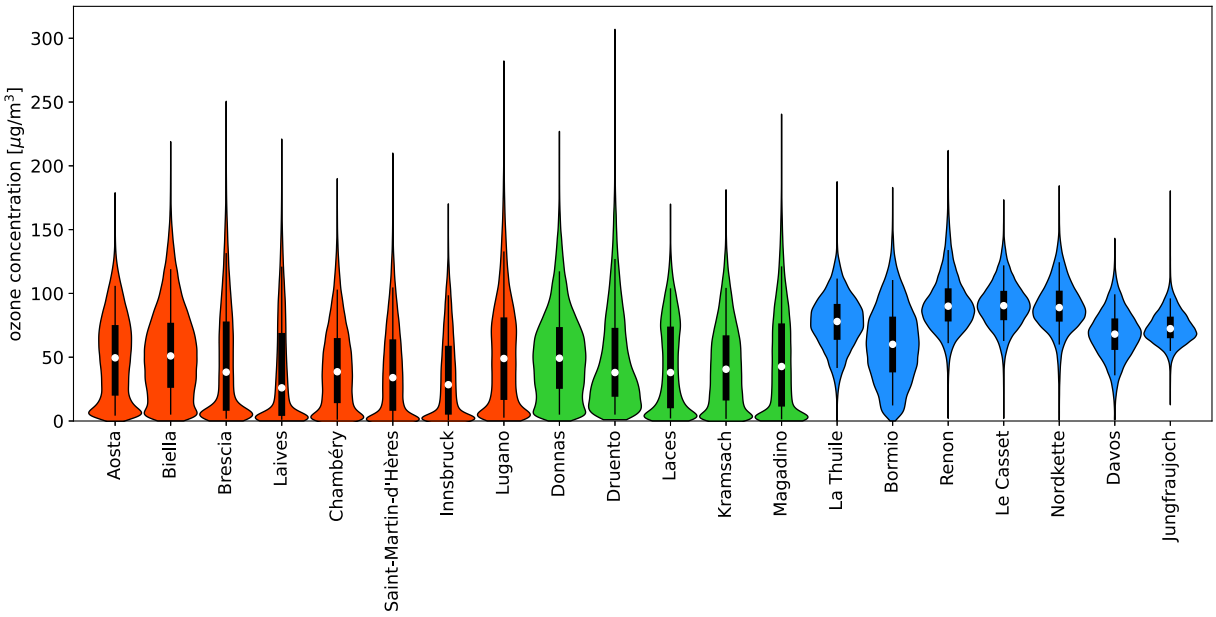


Fig. 4. Distribution of the hourly ozone data for urban (red), rural (green), and mountain (blue) stations. The white dots indicate the median, the vertical bars cover the range between 25 and 75 percentile, and the vertical lines are between 5 and 95 percentile.

cycles. At Jungfraujoch, for instance, the average hourly concentration is practically constant at about 70 μg m^{-3} during all the hours of the day and summer values are just about 20% higher than winter ones. In Lugano, on the contrary, summer average values are around 100 μg m^{-3} while the winter concentration is almost zero.

3.2. The forecasting task

Forecasting the future ozone evolution requires identifying a predictive model that processes the past dynamics, $y(t-d+1), \dots, y(t-1), y(t)$ and estimates the following value $\hat{y}(t+1)$ (one-step forecasting) or the sequence of h future values, $\hat{y}(t+1), \hat{y}(t+2), \dots, \hat{y}(t+h)$ (multi-step forecasting). In this paper, we aim to forecast the hourly ozone concentration for the next day (h equal to 24 h), feeding into the model the values of the past two days (d equal to 48 h). We use only current and lagged O_3 concentrations as inputs without adding any exogenous variable (following the experimental setup presented in Gómez-Losada et al., 2018).

Among the models available for this aim, we adopt artificial neural networks, machine learning tools that have been extensively used for time series forecasting (Tealab, 2018; Nielsen, 2019; Lim and Zohren, 2021; Sangiorgio et al., 2022). Various neural architectures can be employed, ranging from the multi-layer perceptrons (characterized by a feed-forward and dense structure) to more advanced ones such as convolutional and recurrent networks. Convolutional nets process information by sliding convolutional filters over the data. Each filter has a number of parameters to be learned representing the neural network weights. Thanks to the sliding process, the number of weights is strongly reduced with respect to a dense structure with the same number of neurons. Although convolutional nets owe their popularity to image processing, they can be used to tackle sequential tasks by using one-dimensional filters (Sen et al., 2019; Cheng et al., 2021; Durairaj and Mohan, 2022).

Conversely, recurrent neural networks are precisely thought to tackle sequential data using an internal hidden state associated with each recurrent neuron. The presence of the internal states is related to the cyclic (recurrent) connections between the neurons that distinguish such architectures from the feed-forward ones (Sangiorgio and Dercole, 2020; Hewamalage et al., 2021). The parameter sharing across the time steps allows for a reduction in the number of parameters if compared to feed-forward and fully connected structures with the same number of neurons.

In this work, advanced, recurrent neurons known as Long Short-Term Memory (LSTM) cells (Hochreiter and Schmidhuber, 1997) are adopted. They have two internal states and three gates that modulate the information flow in and out of the cell. LSTM neural nets are more robust than naive recurrent nets against vanishing gradients and more efficient in retaining information over extended time intervals. They have been proven to outperform feed-forward and recurrent predictors on several univariate time series forecasting tasks.

The ozone time series are organized into pairs of input and output sequences so that the network can be trained following the traditional supervised learning approach (Taieb et al., 2012; Sangiorgio et al., 2021; Masini et al., 2023). Additionally, the training is performed using the closed-loop (*i.e.*, without teacher forcing) algorithm presented in Sangiorgio and Dercole (2020).

Starting from a null initial internal state ($s(t-d)$) of the LSTM model $f(\cdot)$ and the time series of past values, $y(t-d+1), \dots, y(t-1), y(t)$, the predictions for time steps from $t+1$ to $t+h$ can be computed using the following equations:

$$\begin{aligned}
\left[\hat{y}(t-d+2), s(t-d+1)\right] &= f\Big(y(t-d+1), s(t-d)\Big)\\
&\dots\\
\left[\hat{y}(t), s(t-1)\right] &= f\Big(y(t-1), s(t-2)\Big)\\
\left[\hat{y}(t+1), s(t)\right] &= f\Big(y(t), s(t-1)\Big)\\
\left[\hat{y}(t+2), s(t+1)\right] &= f\Big(\hat{y}(t+1), s(t)\Big)\\
&\dots\\
\left[\hat{y}(t+h), s(t+h-1)\right] &= f\Big(\hat{y}(t+h-1), s(t+h-2)\Big)
\end{aligned}$$

Note that the predicted values from the initial step to time t (in red) are discarded since they refer to the warm-up period, which is necessary for the LSTM to learn the context.

The network is trained using Adam (adaptive momentum estimation) optimizer, with the gradients computed by the traditional backpropagation through time algorithm. Mean squared error (averaged on the 24-hour horizon) is used as the loss function. To allow

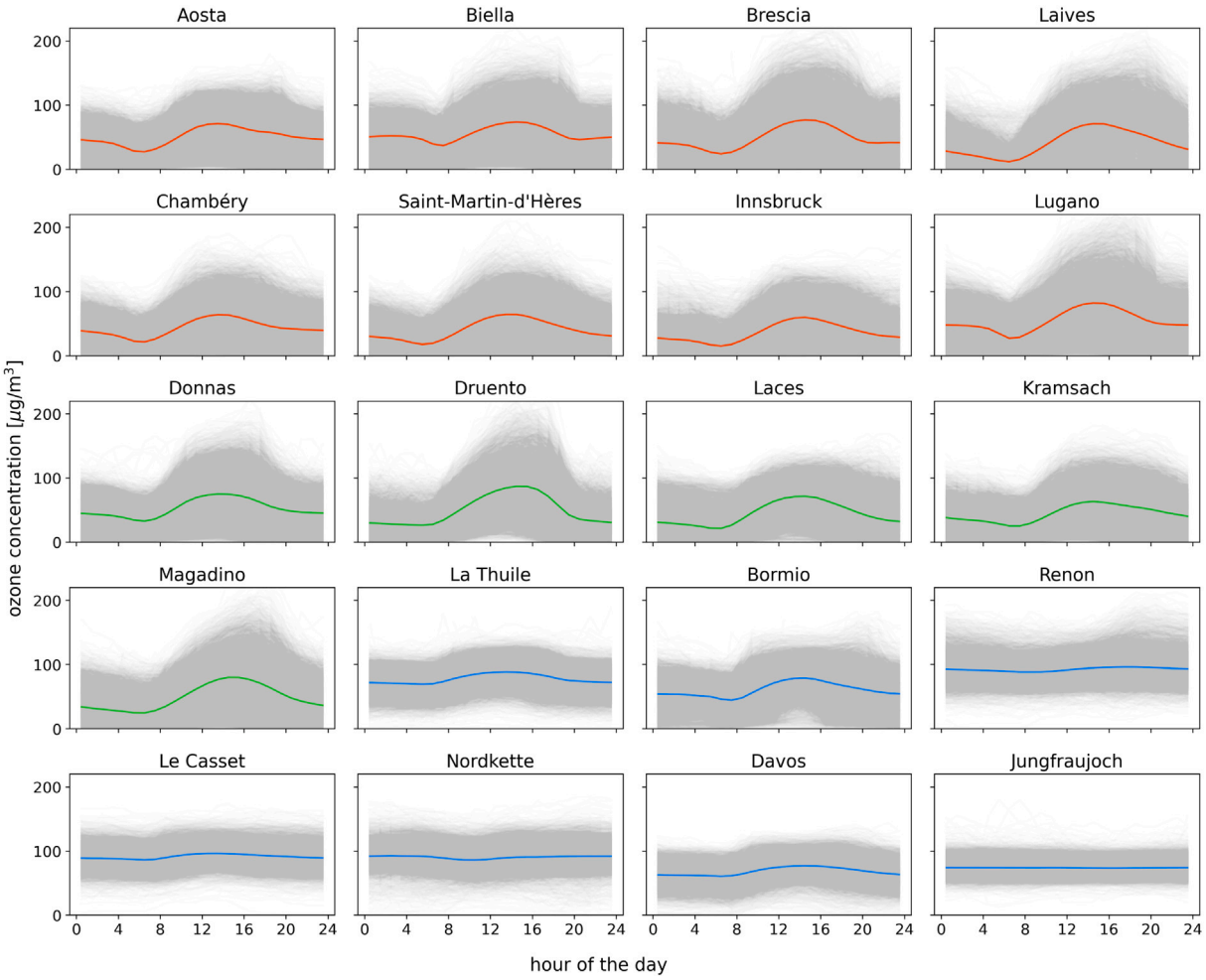


Fig. 5. Average daily cycle for each monitoring station (urban-red, rural-green, mountain-blue). Hourly values are reported in grey.

for a meaningful comparison at an acceptable computational cost, the same neural structure is used for all the sites: two LSTM layers with a fully connected one on top of them. Different complexities of the neural predictors, as well as values of the hyperparameters that regulate the training procedure, have been accounted for. We considered a number of neurons in each hidden layer of 5 or 10; a learning rate (*i.e.*, the step size of the gradient-based optimization) of 10^{-2} or 10^{-3}; a decay of the learning rate of 10^{-3} or 10^{-4}; and a batch size (the number of samples used to perform one iteration of the algorithm) of 256 or 512. Two training runs, starting from different random seeds, are performed for each combination of the hyperparameters to avoid dependence on particularly unlucky initialization. An exhaustive grid search selects the best combination of the hyperparameters for each site.

4. Results and discussion

Table 2 presents the transferability matrix containing the R^2-scores computed on a 24-step forecasting horizon. A predictive model is identified (*i.e.*, trained, validated and tested) for each location on the matrix's rows and then applied, without retraining, to all the locations on the matrix's columns. For instance, the first row of the matrix reports the results obtained with the model identified on Aosta data, the second on Biella data, and so on, as schematically shown in Fig. 1. Similarly, the first column shows the R^2-scores computed by applying the other predictors to Aosta's data, the second column to Biella, and so on. The values on the diagonal thus represent the R^2-scores computed on the identification domain as it traditionally happens in machine learning.

The transferability matrix is not symmetric, meaning that a predictive model identified on site X and applied to site Y performs differently from a model identified for site Y and applied to site X. Another apparently strange feature of the performance matrix is that the values on the diagonal are not always the maximum of their row. This means a model developed for a certain site performs better when applied to another site. Both these features can be explained by looking again at Fig. 5, which shows that the concentration variability of some sites is smaller than in other sites, so a model developed for a site with high variability performs differently, and possibly even better, when applied to a smoother dynamic.

The most evident feature of Table 2 is that it appears partitioned into two areas. The upper left area represents the urban and rural stations where all the R^2 values are of the order of 0.6 and above. On the contrary, the lower right area has many values around 0.3 and below. This result is partly justified by the strong presence of the daily variability at low elevation. As already shown in Figs. 4 and 5, ozone concentration is often close to zero during nighttime. Forecasting in these conditions is thus easier and errors are low in comparison to the high variance of the series. This corresponds to high performance, as measured in terms of R^2. In the mountains, the situation is reversed: values are always high with a much lower variance. The two rectangles showing the results of applying urban and rural models to mountain sites mainly contain very low and even negative values. Davos, and, even more, Bormio show performances somehow like rural sites for being subject to intense tourist flow, as already explained.

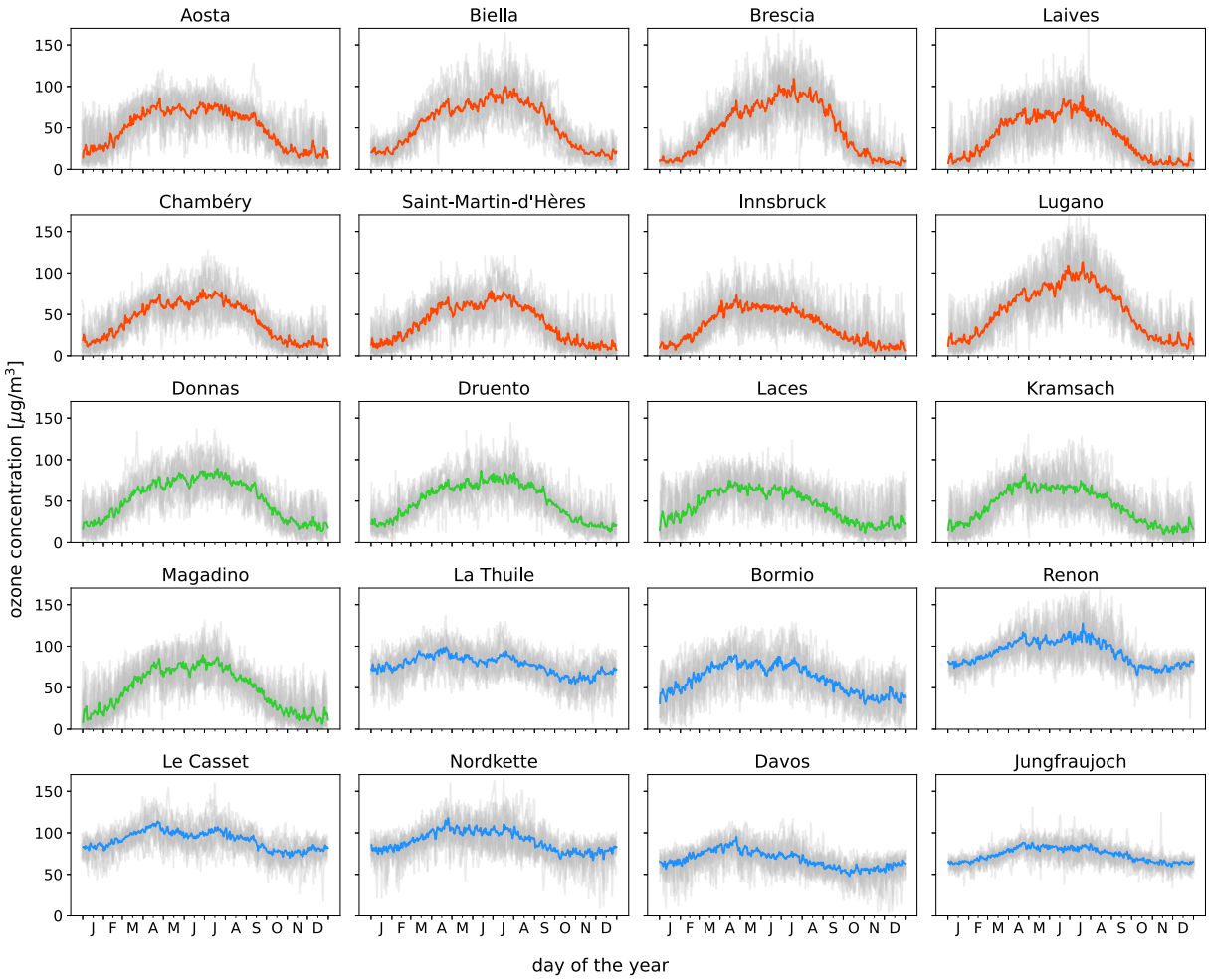


Fig. 6. Average annual cycle for each monitoring station (urban-red, rural-green, mountain-blue). Daily values are reported in grey.

The table can be used to compute the (global) transferability indices for the case at hand. Averaging all the values on the diagonal (self-tests) and all the other values in the table (cross-tests), one obtains the expected performance of using one of the predictors on a new station: the 24-hours R^2 value is between 0.70 and 0.19. The distributions of all the performance values are presented in the upper part of Fig. 9. If we assess transferability using, for instance, the median instead of the average of the two distributions, the indicator range is 0.70 to 0.51. The much higher value of the lower limit is due to the long left tail of the cross-test distribution that strongly reduces the average.

The numerical range obtained depends on the performance metric adopted. For instance, R^2 values computed only on the first six hours of the forecasting horizon are much higher, reflecting the higher precision of the initial forecast. The average performance of urban and rural stations is 0.84, while that of mountain stations is 0.78. It is worth noting that the R^2 increase of some mountain stations is larger than that of other types of stations so the differences among the stations' performances are less evident. This can be explained by looking at the average prediction errors reported in Fig. 7 for three sample cases. In the cases of Chambéry (urban) or Druento (rural), the forecasting error increases in the first eight to ten steps but then reaches a plateau and does not significantly change for longer forecasting horizons. On the contrary, at Nordkette, the error keeps increasing, and the R^2 keeps decreasing with the horizon length. Using the 6-hour values, the transferability indicators are 0.83 to 0.56.

Table 3 shows again the transferability matrix under a different perspective: performances are measured in terms of mean absolute error (MAE). The numerical values are thus very different from previous ones, and their distribution appears also different. In this case, the best values (closest to zero) are mainly concentrated in the lower right corner of the table, meaning that some mountain models perform well (when evaluated with MAE) when applied to other stations of the same type. This can be explained by looking again at Figs. 6 and 5. The yearly and daily variability of these stations is much lower than others. So, even if following their evolution is more difficult (low R^2 values), predictions do not differ too much from observed values (i.e., low MAE). In summary, MAE performance represents a complementary point of view about the accuracy of forecast models, but the transferability evaluation can be equally performed in this case. Again, one can compute the performance of each model on the data of the corresponding station, on those of the other stations, and of the other models on the data of the station of interest. The procedure produces the range 11.6 to 18.3.

Given the presence of several source domains with heterogeneous behaviours, we apply a clustering algorithm to highlight each group's specific features and narrow the expected performance range. A wide variety of clustering algorithms are available (Rokach and Maimon, 2005), and all have a certain degree of subjectivity related, for instance, to the number of clusters to be selected. The algorithm designed for this study takes the transferability matrix as input and groups the sites into a limited number of clusters, each associated with an intra-cluster transferability sub-matrix. The algorithm aggregates the sites so that the variability of the off-diagonal elements of each sub-matrix is minimized. This requires setting up an optimization problem where the objective function is $J = \sum_{k=1}^{K} J_k$, being K the number of clusters and J_k the sum of the squared deviations from the average of the

Table 2
Transferability matrix (R^2-score) of the ozone experiment. The generic cell (i, j) reports the performance of the predictor trained on the ith site and applied on the jth site.

identification \ forecasting	Aosta	Biella	Brescia	Laives	Chambéry	Saint-Martin-d'Hères	Innsbruck	Lugano	Donnas	Druento	Laces	Kramsach	Magadino	La Thuile	Bormio	Renon	Le Casset (Le Monêtier-les-Bains)	Nordkette (Innsbruck)	Davos	Jungfraujoch
Aosta	0.72	0.76	0.75	0.65	0.71	0.63	0.57	0.74	0.71	0.61	0.54	0.67	0.66	0.46	0.60	0.29	0.39	0.24	0.48	0.26
Biella	0.67	0.82	0.74	0.56	0.68	0.52	0.47	0.74	0.73	0.55	0.49	0.63	0.59	0.47	0.60	-0.04	0.05	-0.15	0.43	-0.13
Brescia	0.65	0.75	0.81	0.69	0.70	0.66	0.63	0.75	0.72	0.72	0.60	0.64	0.69	-0.09	0.48	-2.17	-2.22	-2.89	-0.67	-4.67
Laives	0.62	0.58	0.73	0.75	0.65	0.71	0.66	0.66	0.60	0.72	0.63	0.60	0.71	-0.11	0.40	-1.60	-1.47	-1.81	-0.27	-2.05
Chambéry	0.70	0.75	0.78	0.70	0.74	0.72	0.67	0.75	0.72	0.65	0.60	0.68	0.69	0.22	0.56	-0.41	-0.29	-0.61	0.17	-1.07
Saint-Martin-d'Hères	0.66	0.69	0.77	0.73	0.70	0.75	0.69	0.72	0.69	0.75	0.64	0.66	0.72	0.00	0.46	-0.97	-0.92	-1.24	-0.24	-2.83
Innsbruck	0.64	0.66	0.73	0.72	0.69	0.72	0.69	0.67	0.66	0.71	0.63	0.64	0.70	0.01	0.46	-1.09	-1.01	-1.31	-0.14	-2.07
Lugano	0.69	0.77	0.78	0.64	0.69	0.60	0.56	0.78	0.72	0.63	0.56	0.63	0.65	0.34	0.57	0.00	0.09	-0.12	0.20	-0.77
Donnas	0.69	0.78	0.75	0.62	0.70	0.58	0.56	0.74	0.75	0.61	0.50	0.67	0.65	0.44	0.60	0.27	0.37	0.27	0.36	-0.64
Druento	0.62	0.65	0.72	0.70	0.66	0.68	0.61	0.66	0.69	0.81	0.62	0.62	0.71	0.15	0.47	-1.51	-1.27	-1.62	0.14	-0.61
Laces	0.66	0.65	0.71	0.73	0.68	0.70	0.67	0.64	0.66	0.73	0.68	0.64	0.71	0.23	0.53	-0.51	-0.31	-0.55	0.17	-1.11
Kramsach	0.70	0.73	0.75	0.71	0.73	0.72	0.67	0.71	0.71	0.64	0.60	0.70	0.69	0.21	0.55	-1.00	-0.91	-1.30	0.11	-1.49
Magadino	0.66	0.70	0.77	0.73	0.69	0.72	0.67	0.72	0.70	0.78	0.65	0.65	0.74	0.12	0.51	-0.72	-0.68	-0.87	-0.29	-3.51
La Thuile	0.39	0.60	0.39	-0.06	0.26	0.04	-0.16	0.55	0.52	0.21	0.12	0.30	0.28	0.53	0.48	0.53	0.64	0.53	0.53	0.49
Bormio	0.69	0.80	0.72	0.59	0.69	0.56	0.49	0.72	0.72	0.59	0.55	0.65	0.62	0.50	0.65	0.39	0.51	0.35	0.50	0.14
Renon	-1.11	-0.50	-1.27	-2.64	-1.81	-2.30	-3.02	-0.59	-0.55	-1.83	-2.16	-1.28	-1.43	0.27	-0.30	0.62	0.69	0.59	0.32	0.46
Le Casset (Le Monêtier-les-Bains)	-0.72	-0.17	-0.57	-1.22	-0.88	-0.91	-1.22	0.10	-0.26	-0.39	-0.40	-0.49	-0.34	0.28	0.19	0.60	0.68	0.57	0.28	0.40
Nordkette (Innsbruck)	-0.24	0.16	-0.55	-1.55	-0.64	-1.11	-1.66	0.10	-0.01	-0.69	-0.86	-0.43	-0.66	0.18	0.09	0.61	0.68	0.59	0.30	0.45
Davos	0.41	0.57	0.45	0.23	0.40	0.34	0.24	0.54	0.51	0.26	0.21	0.45	0.42	0.52	0.52	0.29	0.43	0.29	0.59	0.47
Jungfraujoch	0.06	0.13	0.26	-0.01	-0.05	-0.06	-0.14	0.27	-0.09	-0.11	-0.03	-0.01	0.09	0.23	0.15	0.35	0.50	0.39	0.34	0.54

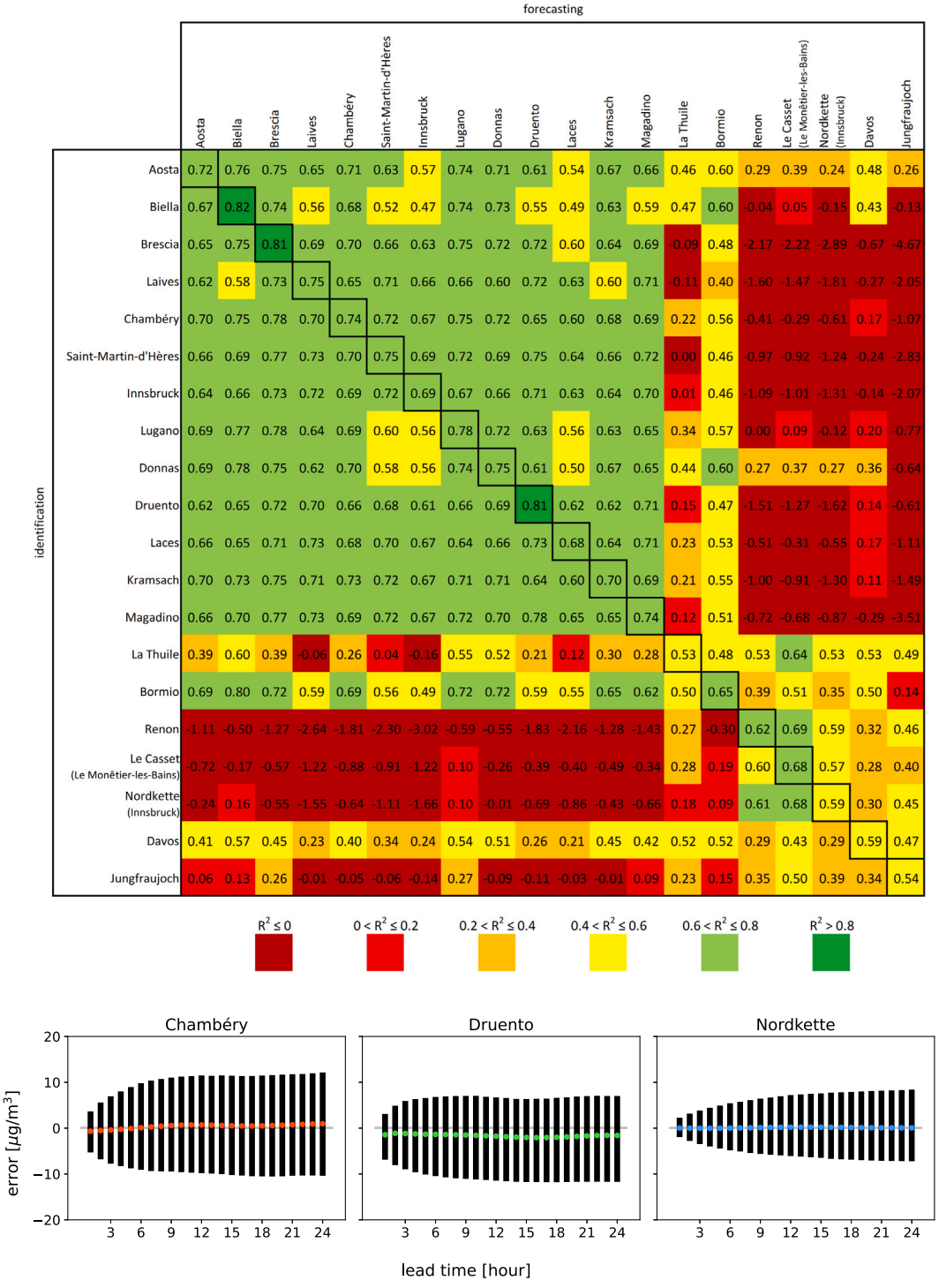


Fig. 7. Box plots of the prediction error for each step of the forecasting horizon (from 1 to 24 h ahead) for Chambéry (urban, in red), Druento (rural, in green), and Nordkette (mountain, in blue). The coloured points indicate the mean, while the vertical bars cover the range between 25 and 75 percentile.

off-diagonal elements for the sub-matrix relative to cluster k. The problem can be solved by considering all the possible combinations (thus adopting a brute force approach) or using a suitable heuristic algorithm for combinatorial optimization (Yang and Zhuang, 2010; Korte et al., 2011) if the exhaustive search requires an unsustainable computational effort.

In the case at hand (with 20 sites), an exhaustive search can be used to obtain the optimal partition for a number of clusters equal to 2 and 3, while the extreme cases of a single cluster and 20 clusters (one for each site) require a single evaluation of the objective function.

Whatever performance indicator (R^2-score or MAE) is selected, the algorithm suggests that the most efficient subdivision is into two clusters (see Fig. 8): the first includes all the urban and rural stations

Table 3
Transferability matrix (MAE) of the ozone experiment. The generic cell (i, j) reports the performance of the predictor trained on the ith site and applied on the jth site.

plus Bormio, and the second includes all the other mountain stations. For instance, when adopting the 24-hour R^2-score, we obtain a transferability range of 0.74 to 0.66 for the first cluster and 0.59 to 0.44 for the second (see Table 4). So, if a new station can be classified in the first cluster for some external reason, the accuracy of one of the existing predictors is expected to be high. The expected performances are lower if it belongs to the second cluster, but the expected range is still limited. If, for some imprecise evaluation, it is assigned to the wrong cluster, the transferability range remains the global one, i.e., evaluated on the entire matrix.

When measuring the performances in terms of MAE, the values are 15.2 to 12.9 for the first cluster and 9.8 to 8.4 for the second (Table 4). As already noted, since the variability of the second cluster is lower, the transferability range is narrower for the second cluster compared to the first.

The analysis reported in Table 4, which only considers the averages, can be extended to the entire distributions as in Fig. 9. Considering, for instance, the case of a single cluster (on top), the blue distribution is that of the principal diagonal elements (self-test, where the source and target domains coincide). This distribution is a reference and represents an upper bound of the accuracy that we can obtain. The light blue distribution, on the other hand, is specific for cross-test (non-diagonal elements). The same analysis can be repeated when two clusters (c1 and c2) are considered.

Table 4
Transferability indicators ranges in terms of R^2-score and MAE.

Transferability indicators range	1 cluster	2 clusters
R^2-score	0.70–0.16	0.74–0.66 0.59–0.44
MAE	11.6–18.3	12.9–15.2 8.4–9.8

To show the practical consequences of these analyses, Fig. 10 reports four examples of the observed versus predicted trajectories. In the first example (top-left panel), we compare a 24-hour time series observed in Brescia (urban) with the predictions of two models trained on Brescia and Lugano (urban) data, respectively. On the opposite side (bottom-right panel), Le Casset data (mountain) are compared with Le Casset and Nordkette (mountain) model predictions. These cases represent what one can expect when a model developed within a cluster is adopted for a station in the same cluster. While being specific episodes, without statistical significance, one can note the high

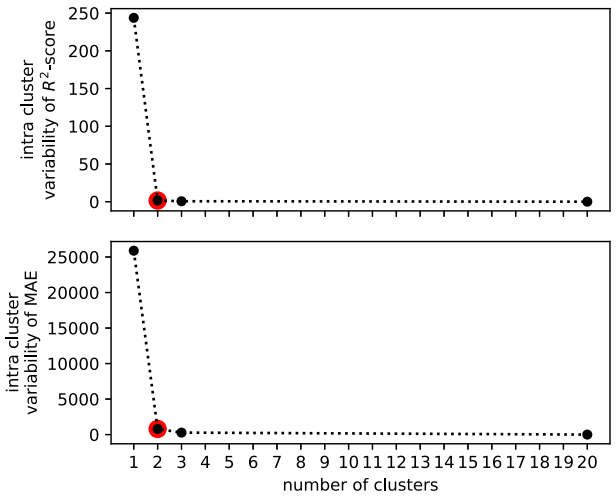


Fig. 8. Intra-cluster variability (sum of the squared deviations from the average of the off-diagonal elements) for different numbers of clusters. The elbow method clearly indicates 2 as the optimal number of clusters.

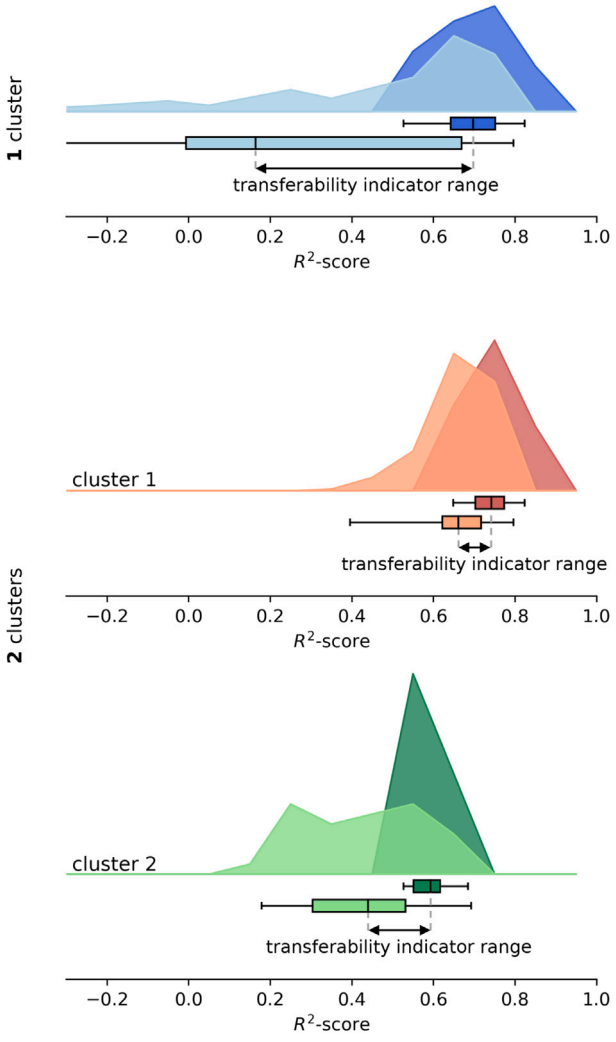


Fig. 9. Distributions of the transferability indicators. The box indicates 25-percentile, average and 75-percentile; whiskers cover the whole distribution range.

accuracy of the predictions that follow quite closely the hourly evolution of the concentration, even 24 h in advance. The other two panels represent the inter-cluster situations. The top-right panel shows the results of forecasting mountain values (Le Casset) using an urban model (Lugano). Vice versa, the bottom-left panel reports the forecasting of urban data (Brescia) with a mountain model (Nordkette). Within each cluster, the loss of accuracy due to using a model developed with other data is quite limited, while the use of a model from the other cluster causes a drop of the predictive accuracy due to the different dynamic characteristic of each cluster. The Lugano model shows some errors in forecasting the constant behaviour typical of mountain sites because it has been trained on data with a strong daily variability. However, following the almost constant values of Le Casset is a relatively simple task and the differences remain limited. On the other hand, the model identified in Nordkette largely departs from Brescia observations since it is unable to reproduce such a periodicity, which is essentially absent (or very limited) in its training data.

5. Conclusion

The problem of the limits of a model domain has been studied since the very beginning of environmental modelling. When dealing with the classical process-driven models, the adaptation to a new context or even a different task could be obtained by modifying some measurable parameters in the model equations. The problem becomes more difficult with the current large diffusion of data-driven models since their parameters have no physical meaning and cannot be measured or even guessed for new sites. It is thus important to be able to evaluate the expected performance of an existing data-driven model to new environmental conditions. Intuitively, performances change the more the new environmental context differs from the original one, but how do we measure and evaluate such differences when sufficient data from the new site are missing? The procedure proposed and tested in this paper indicates the performances one can expect when using a model developed elsewhere for a new site.

The procedure can be adopted for different performance indexes and different clustering algorithms. It can also be extended to models using external inputs besides the autoregressive terms, provided such inputs are equally available in all the considered sites. In the specific case of ozone, we can add inputs about meteorological conditions, particularly wind speed and direction, as often done in the literature. Adding these inputs, we expect an improvement in the performance of each station when training with its own data (the diagonal of the transferability matrix) because the model can take advantage of any meteorological information relevant to the ozone dynamic. However, these relationships are usually quite site-specific, and this may lead to a limited improvement (or even a reduction) of the cross-testing accuracies (off-diagonal elements). As a consequence, we expect, at least for the specific case of ozone considered here, a widening of the transferability ranges. Future studies may address this problem in detail. As to clustering, the algorithms are effective if they can be guided and the results can be interpreted by looking at some additional features not included in the dataset. For instance, in the case of ozone forecasting, the station elevation and the local population density are two characteristics that distinguish the stations into two clear groups. In this respect, the classical subdivision into urban and rural sites does not seem relevant when developing ozone forecasting models. This is possibly due to the duration of the ozone formation process, which means that the air masses transporting the precursors' emissions may move kilometres away from their sources. This phenomenon may be even stronger in the Alpine regions because of the periodic presence of mountain and valley breezes. The results have also shown that, as far as ozone predictions are concerned, a mountain site can be similar to low-altitude environments when characterized by a relevant tourist movement.

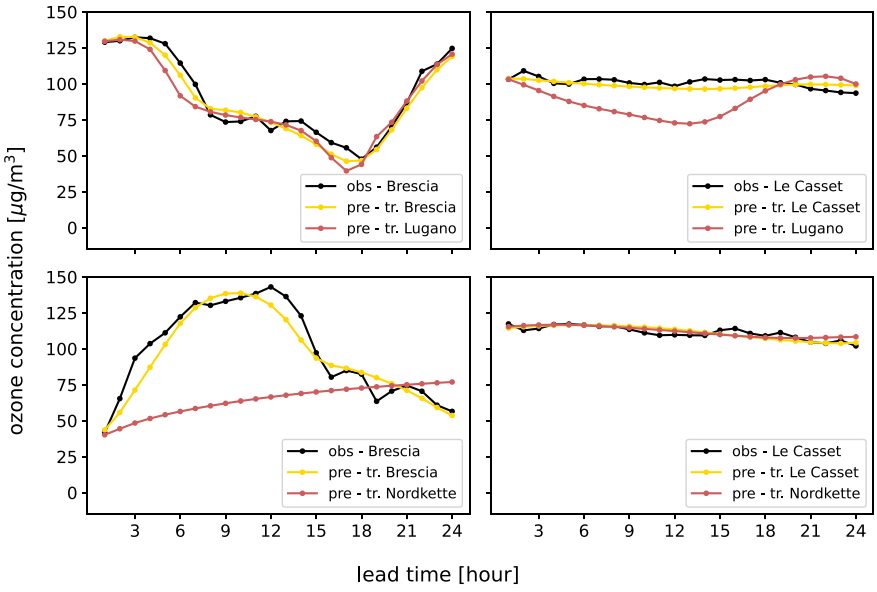


Fig. 10. Four examples of predicted trajectories: Brescia (urban-cluster 1), with predictors identified for Brescia itself and Lugano (urban-cluster 1), from 2021/04/23 02:00 pm (top-left); Le Casset (mountain-cluster 2), with predictors identified for Le Casset itself and Lugano, from 2019/01/02 00:00 (top-right); Brescia, with predictors identified for Brescia itself and Nordkette (mountain-cluster 2), from 2020/06/26 07:00 am (bottom-left); Le Casset, with predictors identified for Le Casset itself and Nordkette, from 2021/04/24 06:00 am (bottom-right).

Overall, the results demonstrate the effectiveness of LSTM models in capturing the complex temporal patterns and nonlinear relationships in ozone concentration data. Multi-step LSTM neural networks have proved to be very precise in forecasting the day-ahead concentrations at hourly time steps in all the tested environments, even without resorting to other meteorological inputs that are sometimes difficult to obtain. This can partly be due to the relatively slow dynamic of the emission changes that generally follow a yearly and weekly pattern that can be easily captured by the structure of the LSTM cells, while the much faster meteorological evolution can be taken into account by the short-term gates.

For the very nature of data-driven models, the model architecture is generally a function of the specific application. For instance, a different architecture could perform better on the stations of a given cluster and also on a specific site. However, given the excellent performance of the LSTM and extensive testing, we expect improvements achievable with ad-hoc network structures to be very limited.

Accurate prediction of ozone concentrations using LSTM models has significant implications for air quality management and public health. The ability to forecast ozone levels with improved accuracy and temporal granularity allows the computation of the standard pollution indicator (the 8-hour running mean), enabling policymakers to implement timely control strategies, issue warnings, and protect vulnerable populations. The findings of this study can contribute to the development of advanced forecasting systems for air pollution monitoring networks in the Alpine Arc and other regions worldwide.

Finally, the proposed transferability assessment procedure is very general and can be applied to different tasks and with different clustering algorithms.

CRediT authorship contribution statement

Matteo Sangiorgio: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Data curation, Conceptualization. **Giorgio Guariso:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors would like to thank G. Brulfert, ATMO Auvergne-Rhône-Alpes, Bron, France, and G. Strickner, Amt der Tiroler Landesregierung, Innsbruck, Austria, for providing data and suggestions.

References

Abdul-Wahab, S.A., Al-Alawi, S.M., 2002. Assessment and prediction of tropospheric ozone concentration levels using artificial neural networks. Environ. Model. Softw. 17 (3), 219–228.

Agirre-Basurko, E., Ibarra-Berastegi, G., Madariaga, I., 2006. Regression and multilayer perceptron-based models to forecast hourly O3 and NO2 levels in the Bilbao area. Environ. Model. Softw. 21 (4), 430–446.

Alijla, B.O., Lim, C.P., Wong, L.-P., Khader, A.T., Al-Betar, M.A., 2018. An ensemble of intelligent water drop algorithm for feature selection optimization problem. Appl. Soft Comput. 65, 531–541.

Betancourt, C., Li, C.W., Kleinert, F., Schultz, M.G., 2023. Graph machine learning for improved imputation of missing tropospheric ozone data. Environ. Sci. Technol..

Biancofiore, F., Verdecchia, M., Di Carlo, P., Tomassetti, B., Aruffo, E., Busilacchio, M., Bianco, S., Di Tommaso, S., Colangeli, C., 2015. Analysis of surface ozone using a recurrent neural network. Sci. Total Environ. 514, 379–387.

Cakaj, A., Qorri, E., Coulibaly, F., De Marco, A., Agathokleous, E., Leca, S., Sicard, P., 2023. Assessing surface ozone risk to human health and forests over time in Poland. Atmos. Environ. 119926.

Camastra, F., Capone, V., Ciaramella, A., Riccio, A., Staiano, A., 2022. Prediction of environmental missing data time series by support vector machine regression and correlation dimension estimation. Environ. Model. Softw. 150, 105343.

Chattopadhyay, G., Chattopadhyay, S., 2008. A probe into the chaotic nature of total ozone time series by correlation dimension method. Soft Comput. 12 (10), 1007–1012.

Chen, J.-L., Islam, S., Biswas, P., 1998. Nonlinear dynamics of hourly ozone concentrations: Nonparametric short term prediction. Atmos. Environ. 32 (11), 1839–1848.

Cheng, W., Wang, Y., Peng, Z., Ren, X., Shuai, Y., Zang, S., Liu, H., Cheng, H., Wu, J., 2021. High-efficiency chaotic time series prediction based on time convolution neural network. Chaos Solitons Fractals 152, 111304.
Cordery, I., 1971. Estimation of design hydrographs for small rural catchments. J. Hydrol. 13, 263–277.
Diao, R., Chao, F., Peng, T., Snooke, N., Shen, Q., 2013. Feature selection inspired classifier ensemble reduction. IEEE Trans. Cybern. 44 (8), 1259–1268.
Durairaj, D.M., Mohan, B.K., 2022. A convolutional neural network based approach to financial time series prediction. Neural Comput. Appl. 34 (16), 13319–13337.
Farahani, A., Voghoei, S., Rasheed, K., Arabnia, H.R., 2021. A brief review of domain adaptation. In: Advances in Data Science and Information Engineering. Springer, pp. 877–894.
Finlayson-Pitts, B., Pitts, Jr., J., 1993. Atmospheric chemistry of tropospheric ozone formation: Scientific and regulatory implications. Air Waste 43 (8), 1091–1100.
Fong, I.H., Li, T., Fong, S., Wong, R.K., Tallon-Ballesteros, A.J., 2020. Predicting concentration levels of air pollutants by transfer learning and recurrent neural network. Knowl.-Based Syst. 192, 105622.
Gauch, M., Mai, J., Lin, J., 2021. The proper care and feeding of CAMELS: How limited training data affects streamflow prediction. Environ. Model. Softw. 135, 104926.
Ghosh, R., Li, B., Tayal, K., Kumar, V., Jia, X., 2022. Meta-transfer learning: An application to streamflow modeling in river-streams. In: 2022 IEEE International Conference on Data Mining. ICDM, IEEE, pp. 161–170.
Glorot, X., Bordes, A., Bengio, Y., 2011. Domain adaptation for large-scale sentiment classification: A deep learning approach. In: ICML.
Glynis, K., Kapelan, Z., Bakker, M., Taormina, R., 2023. Leveraging transfer learning in LSTM neural networks for data-efficient burst detection in water distribution systems. Water Resour. Manag. 1–20.
Gómez-Losada, Á., Asencio-Cortés, G., Martínez-Álvarez, F., Riquelme, J.C., 2018. A novel approach to forecast urban surface-level ozone considering heterogeneous locations and limited information. Environ. Model. Softw. 110, 52–61.
Guariso, G., Nunnari, G., Sangiorgio, M., 2020. Multi-step solar irradiance forecasting and domain adaptation of deep neural networks. Energies 13 (15), 3987.
Guerreiro, C.B., Foltescu, V., De Leeuw, F., 2014. Air quality status and trends in Europe. Atmos. Environ. 98, 376–384.
Hewamalage, H., Bergmeir, C., Bandara, K., 2021. Recurrent neural networks for time series forecasting: Current status and future directions. Int. J. Forecast. 37 (1), 388–427.
Himeur, Y., Elnour, M., Fadli, F., Meskin, N., Petri, I., Rezgui, Y., Bensaali, F., Amira, A., 2022. Next-generation energy systems for sustainable smart cities: Roles of transfer learning. Sustainable Cities Soc. 104059.
Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.
Kerimov, B., Bentivoglio, R., Garzón, A., Isufi, E., Tscheikner-Gratl, F., Steffelbauer, D.B., Taormina, R., 2023. Assessing the performances and transferability of graph neural network metamodels for water distribution systems. J. Hydroinform. jh2023031.
Korte, B.H., Vygen, J., Korte, B., Vygen, J., 2011. Combinatorial Optimization, vol. 1, Springer.
Lee, E.W.M., Lim, C.P., Yuen, R.K.K., Lo, S., 2004. A hybrid neural network model for noisy data regression. IEEE Trans. Syst. Man Cybern. B 34 (2), 951–960.
Lelieveld, J., Dentener, F.J., 2000. What controls tropospheric ozone? J. Geophys. Res.: Atmos. 105 (D3), 3531–3551.
Lim, B., Zohren, S., 2021. Time-series forecasting with deep learning: A survey. Phil. Trans. R. Soc. A 379 (2194), 20200209.
Lin, X., Trainer, M., Liu, S., 1988. On the nonlinearity of the tropospheric ozone production. J. Geophys. Res.: Atmos. 93 (D12), 15879–15888.
Long, M., Cao, Y., Wang, J., Jordan, M., 2015. Learning transferable features with deep adaptation networks. In: International Conference on Machine Learning. PMLR, pp. 97–105.
Ma, J., Cheng, J.C., Lin, C., Tan, Y., Zhang, J., 2019. Improving air quality prediction accuracy at larger temporal resolutions using deep learning and transfer learning techniques. Atmos. Environ. 214, 116885.
Ma, W., Yuan, Z., Lau, A.K., Wang, L., Liao, C., Zhang, Y., 2022. Optimized neural network for daily-scale ozone prediction based on transfer learning. Sci. Total Environ. 827, 154279.
Maciąg, P.S., Kasabov, N., Kryszkiewicz, M., Bembenik, R., 2019. Air pollution prediction with clustering-based ensemble of evolving spiking neural networks and a case study for London area. Environ. Model. Softw. 118, 262–280.
Masini, R.P., Medeiros, M.C., Mendes, E.F., 2023. Machine learning advances for time series forecasting. J. Econ. Surv. 37 (1), 76–111.
Masood, A., Ahmad, K., 2021. A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance. J. Clean. Prod. 322, 129072.
Méndez, M., Merayo, M.G., Núñez, M., 2023. Machine learning algorithms to forecast air quality: a survey. Artif. Intell. Rev. 56 (9), 10031–10066.
Nielsen, A., 2019. Practical Time Series Analysis: Prediction with Statistics and Machine Learning. O'Reilly Media.
O'Donnell, T., 1976. Special techniques of hydrological systems modelling. Hydrol. Sci. J. 21 (1), 17–25.
Pan, S.J., Yang, Q., 2009. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22 (10), 1345–1359.
Paoletti, E., De Marco, A., Racalbuto, S., 2007. Why should we calculate complex indices of ozone exposure? Results from mediterranean background sites. Environ. Monit. Assess. 128, 19–30.
Peng, L., Wu, H., Gao, M., Yi, H., Xiong, Q., Yang, L., Cheng, S., 2022. TLT: Recurrent fine-tuning transfer learning for water quality long-term prediction. Water Res. 225, 119171.
Pisoni, E., Farina, M., Carnevale, C., Piroddi, L., 2009. Forecasting peak air pollution levels using NARX models. Eng. Appl. Artif. Intell. 22 (4–5), 593–602.
Priyatikanto, R., Lu, Y., Dash, J., Sheffield, J., 2023. Improving generalisability and transferability of machine-learning-based maize yield prediction model through domain adaptation. Agricult. Forest Meteorol. 341, 109652.
Rokach, L., Maimon, O., 2005. Clustering methods. In: Data Mining and Knowledge Discovery Handbook. Springer, pp. 321–352.
Sangiorgio, M., Dercole, F., 2020. Robustness of LSTM neural networks for multi-step forecasting of chaotic time series. Chaos Solitons Fractals 139, 110045.
Sangiorgio, M., Dercole, F., Guariso, G., 2021. Forecasting of noisy chaotic systems with deep neural networks. Chaos Solitons Fractals 153, 111570.
Sangiorgio, M., Dercole, F., Guariso, G., 2022. Neural approaches for time series forecasting. In: Deep Learning in Multi-Step Prediction of Chaotic Dynamics: from Deterministic Models to Real-World Systems. Springer, pp. 43–57.
Sangiorgio, M., Guariso, G., 2023. Deep neural network adaptation to different environmental contexts: A case study of ozone forecast. IFAC-PapersOnLine 56 (2), 8290–8295.
Sen, R., Yu, H.-F., Dhillon, I.S., 2019. Think globally, act locally: A deep neural network approach to high-dimensional time series forecasting. In: Advances in Neural Information Processing Systems, vol. 32.
Sfetsos, A., Siriopoulos, C., 2004. Time series forecasting with a hybrid clustering scheme and pattern recognition. IEEE Trans. Syst. Man Cybern.-A 34 (3), 399–405.
Singh, V., Carnevale, C., Finzi, G., Pisoni, E., Volta, M., 2011. A cokriging based approach to reconstruct air pollution maps, processing measurement station concentrations and deterministic model simulations. Environ. Model. Softw. 26 (6), 778–786.
Stowell, J.D., min Kim, Y., Gao, Y., Fu, J.S., Chang, H.H., Liu, Y., 2017. The impact of climate change and emissions control on future ozone levels: Implications for human health. Environ. Int. 108, 41–50.
Tai, A.P., Martin, M.V., Heald, C.L., 2014. Threat to future global food security from climate change and ozone air pollution. Nature Clim. Change 4 (9), 817–821.
Taieb, S.B., Bontempi, G., Atiya, A.F., Sorjamaa, A., 2012. A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. Expert Syst. Appl. 39 (8), 7067–7083.
Tealab, A., 2018. Time series forecasting using artificial neural networks methodologies: A systematic review. Future Comput. Inform. J. 3 (2), 334–340.
US EPA, 2020. Integrated Science Assessment (ISA) for Ozone and Related Photochemical Oxidants (Final Report, Apr 2020). US Environmental Protection Agency Washington, DC, USA.
Wang, P., Chen, Y., Hu, J., Zhang, H., Ying, Q., 2018. Attribution of tropospheric ozone to NOx and VOC emissions: Considering ozone formation in the transition regime. Environ. Sci. Technol. 53 (3), 1404–1412.
Wang, W.-N., Cheng, T.-H., Gu, X.-F., Chen, H., Guo, H., Wang, Y., Bao, F.-W., Shi, S.-Y., Xu, B.-R., Zuo, X., et al., 2017. Assessing spatial and temporal patterns of observed ground-level ozone in China. Sci. Rep. 7 (1), 3651.
Willard, J.D., Read, J.S., Appling, A.P., Oliver, S.K., Jia, X., Kumar, V., 2021. Predicting water temperature dynamics of unmonitored lakes with meta-transfer learning. Water Resour. Res. 57 (7), e2021WR029579.
Xu, Y., Lin, K., Hu, C., Wang, S., Wu, Q., Zhang, L., Ran, G., 2023. Deep transfer learning based on transformer for flood forecasting in data-sparse basins. J. Hydrol. 625, 129956.
Yang, M., Yang, Q., Shao, J., Wang, G., Zhang, W., 2023. A new few-shot learning model for runoff prediction: Demonstration in two data scarce regions. Environ. Model. Softw. 162, 105659.
Yang, J., Zhuang, Y., 2010. An improved ant colony optimization algorithm for solving a complex combinatorial optimization problem. Appl. Soft Comput. 10 (2), 653–660.
Yao, S., Kang, Q., Zhou, M., Rawa, M.J., Albeshri, A., 2022. Discriminative manifold distribution alignment for domain adaptation. IEEE Trans. Syst. Man Cybern.: Syst. 53 (2), 1183–1197.
Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? In: Advances in Neural Information Processing Systems, vol. 27.
You, K., Long, M., Cao, Z., Wang, J., Jordan, M.I., 2019. Universal domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2720–2729.
Yuan, Z., Liu, J., Liu, Y., Zhang, Q., Li, Y., Li, Z., 2022. A two-stage modelling method for multi-station daily water level prediction. Environ. Model. Softw. 156, 105468.
Zhang, J., Wei, Y., Fang, Z., 2019. Ozone pollution: A major health hazard worldwide. Front. Immunol. 10, 2518.
Zhang, B., Zhang, Y., Jiang, X., 2022. Feature selection for global tropospheric ozone prediction based on the BO-XGBoost-RFE algorithm. Sci. Rep. 12 (1), 9244.
Zolghadri, A., Monsion, M., Henry, D., Marchionini, C., Petrique, O., 2004. Development of an operational model-based warning system for tropospheric ozone concentrations in Bordeaux, France. Environ. Model. Softw. 19 (4), 369–382.