Article

A Wavenet-Based Virtual Sensor for PM_{10} Monitoring

Claudio Carnevale *, Enrico Turrini [ID]**, Roberta Zeziola, Elena De Angelis** [ID] **and Marialuisa Volta**

Department of Mechanical and Industrial Engineering, University of Brescia, I-25123 Brescia, Italy;
enrico.turrini@unibs.it (E.T.); r.zeziola001@unibs.it (R.Z.); e.deangelis@unibs.it (E.D.A.);
marialuisa.volta@unibs.it (M.V.)
* Correspondence: claudio.carnevale@unibs.it

Abstract: In this work, a virtual sensor for PM_{10} concentration monitoring is presented. The sensor is based on wavenet models and uses daily mean NO_2 concentration and meteorological variables (wind speed and rainfall) as input. The methodology has been applied to the reconstruction of PM_{10} levels measured from 14 monitoring stations in Lombardy region (Italy). This region, usually affected by high levels of PM_{10}, is a challenging benchmarking area for the implemented sensors. Neverthless, the performances are good with relatively low bias and high correlation.

Keywords: virtual sensors; wavenet; air quality

1. Introduction

Exposure to high levels of particulate matter (PM_{10}) is a big social problem [1] due to its impacts on human health, with effects including pulmonary and cardio-vascular diseases [2,3]. One of the main challenges in decision making related to PM_{10} control is that, usually, win–win solutions that also consider other pollutants, such as nitrogen oxides (NO_2) and ozone (O_3), are complex to identify and implement [4–7]. For this reason, having detailed information about the level of all of the significant air pollutants over a certain area is a key issue in decision-making processes. In this context, the use of integrated information coming from regional networks and novel/private networks supported by low-cost technology [8,9] has become more and more important, which has been mainly due to the fact that they can provide suitable information for chemical transport models (CTMs), allowing them to compute concentrations far away from the official monitoring network stations [10–12].

In principle, four main techniques for the measurement of PM_{10} are presented in literature [13]: (1) gravimetric analysis of pumped and filtered particles; (2) tapering element oscillating microbalance (TEOM); (3) beta-attenuation; (4) light scattering. The first three of these techniques are quite expensive, so their use is limited to regional authorities, private companies and research groups [13]. Light scattering, instead, is a relatively low-cost technique, but it is often affected by consistent biases [14].

The objective of this work is to evaluate the possibility of implementing a virtual sensor for PM_{10} daily mean concentration starting from the data measured by sensors detecting other pollutants and meteorological variables. In particular, the virtual sensors applied in this work are based on NO_2 daily mean concentration and meteorological variables, such as wind speed, rainfall, relative humidity and temperature.

As indicated by the name, virtual sensors can be broadly described as a software that allows us to compute the value of a certain variable without direct measurement considering measurements that are physically/chemically related to the variable that should be reproduced [15]. They assume a key role when it is not possible to place a physical sensor due to any kind of limitations (e.g., unreachable position, high cost). There are two possible approaches to virtual sensor implementation:

1. Data-driven: in this approach, time series of input and output variables are collected from direct measurement and are used to compute a mathematical, approximated relationship between the measured variables' and sensors' output [16];
2. Deterministic: in this approach, the (eventually approximated) physical/chemical relationships among input and output variables are used to compute the unmeasured variable through the virtual sensor [17].

This work presents a data-driven approach based on wavenet models to implement a PM_{10} virtual sensor using NO_2 and meteorological variables. All these variables are strictly related to the phenomena involved in the formation and accumulation of PM_{10} in atmosphere; their choice is due to the presence in the literature of low-cost sensors with performances that are adequate [18] enough to identify a virtual sensor, therefore allowing the definition of a low-cost PM_{10} measuring network. Wavenets are data-driven models resulting from the integration of wavelet theory and neural network models [19]. Their main applications are related to sound management/filtering [20], even if their nonlinear function approximation (and thus forecasting) properties have been applied with good results also in other fields such as energy systems [19,21]. These approximation properties make them suitable for environmental monitoring and forecasting applications, but still, there is no literature related to their application to reproduce PM_{10} or other air quality pollutants. Therefore, since artificial neural networks are widely used in this field [4,22,23], wavenets could also be useful for the definition of a PM_{10} virtual sensor. The paper is organized in two main parts, a methodological one (Section 2) where the basics of the artificial neural network, wavelet theory and wavenets are introduced and a second part presenting the evaluation of the results on a test case.

2. Materials and Methods

In this section, the theoretical framework used to derive a virtual air quality sensor based on wavenets [24] is presented.

2.1. Artificial Neural Networks

Artificial Neural Networks (ANNs) are functions approximating human brain behavior, considered as a network of smaller units, called neurons, representing the information processing unit (Figure 1).
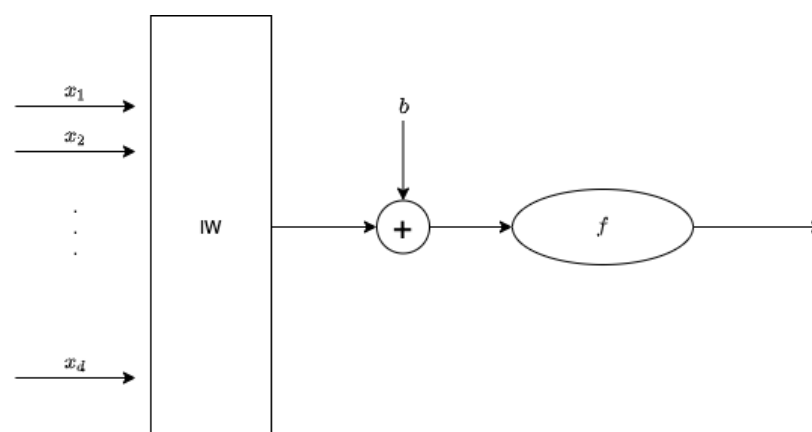

Figure 1. Typical neuron model.

Each input x_i of the network is multiplied by a corresponding weight w_i, analogous to a synaptic force; then all the weighted inputs are added together, including also a bias b term in order to compute the activation level x of the neuron. The output signal $y(x)$ is

usually a nonlinear function $f(x)$ of the activation level. Hence, the typical neuron model is represented as (1):

$$y(x) = f\left(\sum_{i=1}^{d} x_i \times w_i + b\right) \tag{1}$$

where d is the length of the input vector.

The approximation capacity of a single neuron is quite limited; to overcome this, they are collected in layers sharing the same input. The final structure of a neural network is obtained by connecting several layers, as in the case of the two-layer feedforward neural network in Figure 2.
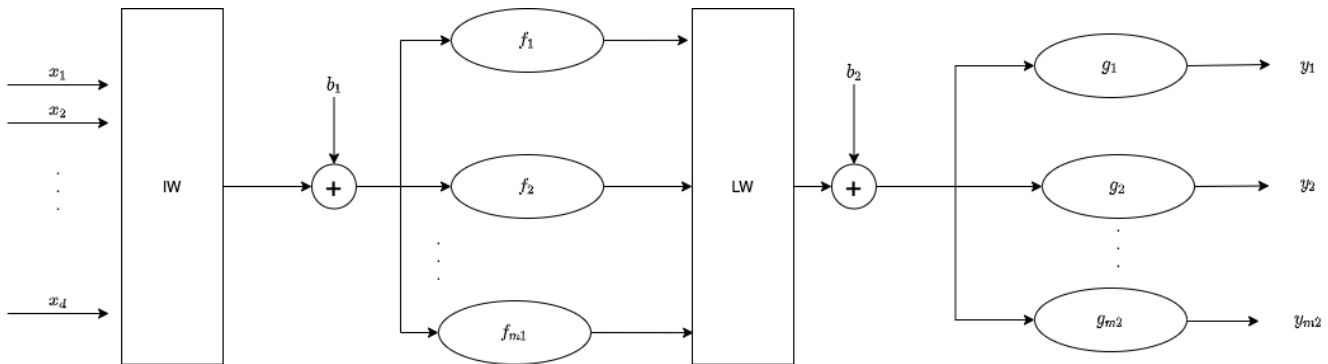

Figure 2. Two-layer feedforward neural network structure.

In this case, the output $y(x)$ can be computed as:

$$y(x) = g(LW \times f(IW \times x + b_1) + b_2) \tag{2}$$

where $y(x) \in \mathbb{R}^{m2}$ is the output of the network, $x \in \mathbb{R}^{d}$ is its input vector, $f : \mathbb{R}^d \to \mathbb{R}^{m1}$ and $g : \mathbb{R}^{m1} \to \mathbb{R}^{m2}$ are the activation functions of the hidden and output layers, respectively, and, finally, $m1$ and $m2$ are the lengths of the activation function output and the neural network output. The bias terms b_1, $b_2 \in \mathbb{R}^{m1}$ and the weight matrices $IW \in \mathbb{R}^{m1 \times d}$ and $LW \in \mathbb{R}^{m2 \times m1}$ are computed during the training phase. Even if the number of layers of an artificial neural network can be higher than 2, following the proof of the Cybenko approximation theorem, and in order to limit the complexity of the network, in real applications only a two layers neural network is used [25].

2.2. Wavelets and Wavenet Models

Wavelets are a family of orthonormal basis functions that can be used to perform transformations among spaces. Their use ranges from function approximation to audio compression [26–28]. The wavelet approximation theory is strictly related to multi-resolution analysis [26]. In this context, a function $h(x)$ can be approximated using the so-called wavelet (mother) and scaling (father) functions, as:

$$h(x) = \sum_{k} c_{j_0}(k)\phi_{j_0,k}(x) + \sum_{j=j_0}^{\infty} \sum_{k} d_j(k)\psi_{j,k}(x) \tag{3}$$

where:

- $c_{j_0}(k)$ are the scaling coefficients;
- $d_j(k)$ are the details (wavelet) coefficient;
- $\phi_{j_0,k}(x)$ is the selected scaling (father) function family;
- $\psi_{j,k}(x)$ is the selected wavelet (mother) function family.

The computation of the scaling and wavelet coefficient is strongly connected to the selected wavelet family (considered as the couple wavelet/scaling functions). Up to now, a number of different functions has been considered and are currently used. More details about wavelet transformation can be found in [26–28].

Wavenets (wavelet networks) [24] can be considered as a one hidden layer network with wavelets as activation functions. In particular, the wavenet output $Y(x)$ for an input $x \in \mathbb{R}^d$ can be computed as:

$$
\begin{aligned}
Y(x) = WN(x) = (x - r)G + \\
+ \sum_{i=1}^{n_s} a_{s\_i} \phi(b_{s\_i}((x - r)Q - c_{s\_i}) + \\
+ \sum_{j=1}^{n_w} a_{w\_j} \psi(b_{w\_j}((x - r)Q - c_{s\_j})
\end{aligned}
\tag{4}
$$

where $\phi(z) = e^{-0.5z \cdot z'}$, $z = b_{s\_i}((x - r)Q - c_{s\_i}$ is the scaling function, $\psi(t) = (m - t \cdot t')e^{-0.5t \cdot t'}$, $t = b_{w\_j}((x - r)Q - c_{s\_j})$ is the wavelet function, $x \in \mathbb{R}^{1 \times d}$ is the row vector input of the wavenet, a_{s_i}, b_{s_i}, a_{w_i}, b_{w_i}, $r \in \mathbb{R}^{1 \times d}$, $G \in \mathbb{R}^{d \times 1}$ and $Q \in \mathbb{R}^{d \times d}$ are the parameters to be computed during the training.

The comparison between Equations (2) and (4) shows that the wavenet can be considered as a neural network with the function:

$$
f(\cdot) = \begin{bmatrix} \phi(\cdot) \\ \psi(\cdot) \end{bmatrix}
\tag{5}
$$

as the activation function of the hidden layer.

When the phenomena to model with the wavenet is dynamical, the wavenet is fed by an input vector $x(t)$ that is the output of a time delay phase:

$$
x(t) = \begin{bmatrix} u_1(t) \\ u_1(t - 1) \\ \dots \\ u_1(t - n_1) \\ \dots \\ u_m(t) \\ u_m(t - 1) \\ \dots \\ u_m(t - n_m) \end{bmatrix}
\tag{6}
$$

where $u_1 \dots u_m$ are the variables selected to compute the output $y(t)$ of the overall system. In this work, since the PM_{10} formation, accumulation and removal are clearly dynamical processes, the system structure presented in Figure 3 is used.
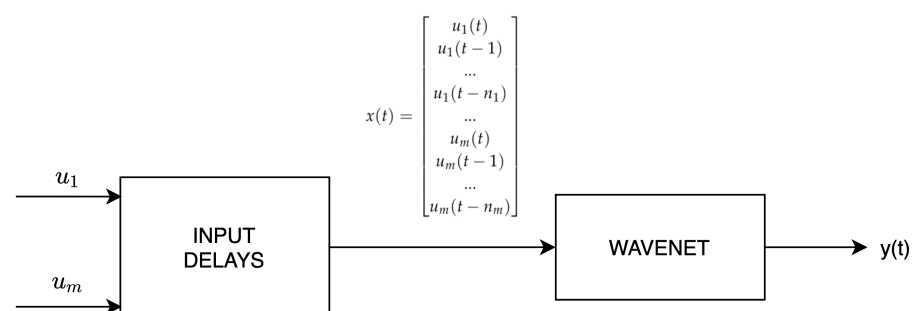

Figure 3. Wavenet structure.

3. Results and Discussion

3.1. Case Study and Dataset Definition

The aim of this work is the definition of a virtual sensor to compute PM_{10} daily average concentrations starting from the measured data of daily average NO_2 concentration and the measured values of two meteorological variables: average daily wind speed WS, total daily rainfall RF, average daily relative humidity RH and average daily temperature T. The selection of NO_2 as the input variable is due to the fact that its levels are strongly related to PM_{10} ones, as they shared some emission drivers (i.e., road traffic, domestic heating) and chemical paths (i.e., formation of secondary inorganic aerosol starting from the ammonium nitrates). On the other hand, the selected meteorological variables can be related to general deposition or dispersion conditions (mainly rainfall and wind speed) or to the formation of secondary aerosol by condensation. Thus, the $Y(x)$ in Equation (4) is the daily PM_{10} concentration computed by the model, which is referred to as $n\hat{PM}_{10}(x)$ from now on. Moreover, the input x of the wavenet function is time dependent, so $x = x(t)$, and it includes both NO_2 concentrations and meteorological variables for the day t and the previous days, as in:

$$x(t) = \begin{bmatrix} NO_2(t) \\ NO_2(t-1) \\ ... \\ NO_2(t-n_{NO_2}) \\ ... \\ WS(t) \\ WS(t-1) \\ ... \\ WS(t-n_{WS}) \\ ... \\ RF(t) \\ RF(t-1) \\ ... \\ RF(t-n_{RF}) \\ ... \\ RH(t) \\ RH(t-1) \\ ... \\ RH(t-n_{RH}) \\ ... \\ T(t) \\ T(t-1) \\ ... \\ T(t-n_T) \end{bmatrix} \tag{7}$$

In order to test the presented methodology, a series of models has been trained and validated to reproduce the PM_{10} daily mean concentrations starting from different input measured by the Lombardy region monitoring network. The work has been tested using data measured by 14 monitoring stations belonging to the Lombardy region (Italy) monitoring network (Figure 4).

More in detail, the data from year 2019 have been used ($365 \times 14 = 5110$ available raw data tuples). The performance evaluation for the different models has been performed using a leave-p-out approach with $p = 4$. Following this approach, 100 tests have been performed for each model configuration, with 10 stations being used for the identification, and the data for $p = 4$ being randomly selected as stations queued in order to define the metastation used for the validation.
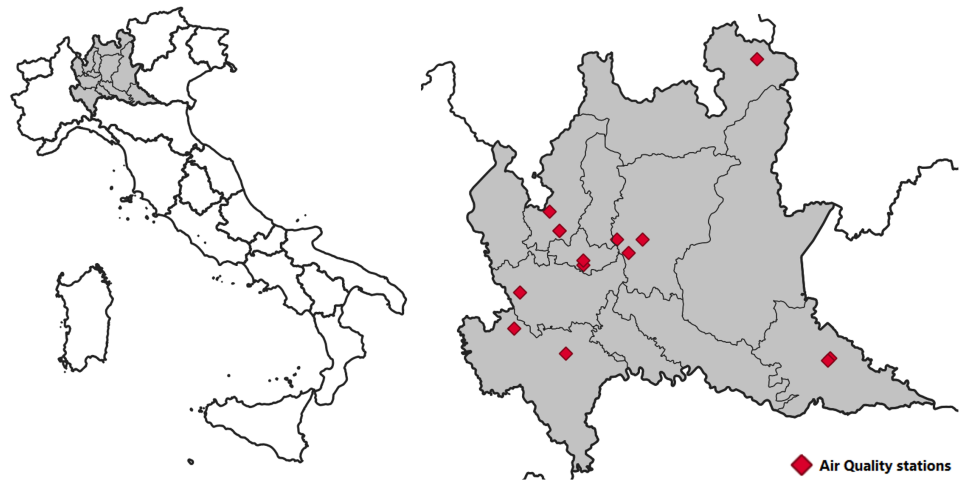
Figure 4. Domain and measuring stations.

3.2. Configuration Tests

In order to evaluate the capability of the methodology presented in Section 2 to compute PM_{10} concentrations, all the possible configurations among the input variables have been considered, and the relative models $PM_{10} = WN(x)$ trained.

In principle, the different configurations can be grouped into three categories:

- Configurations including only NO_2 concentration as input;
- Configurations including only meteorological variables as input;
- Configurations including both NO_2 concentrations and meteorological variables as input.

For each test, an analysis of the memory of the systems, i.e., an evaluation of the performances of varying n_{NO_2}, n_{WS}, n_{RF}, n_{RH} and n_T, has been performed. On the basis of the knowledge of the phenomena related to the formation of PM_{10} in atmosphere, a maximum value of 5 days can be considered for these parameters. Each model has been evaluated on the basis of the following three different statistical indexes:

- Normalized Root Mean Squared Deviation:

$$NRMSD = \frac{\sqrt{\frac{\sum_{t=1}^{T}(PM_{10}(t)-P\hat{M}_{10}(t))^2}{n}}}{PM_{10}^{max}-PM_{10}^{min}}$$

- Root Mean Squared Error

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(PM_{10}(t)-P\hat{M}_{10}(t))^2}{T}}$$

- Correlation Coefficient

$$Correlation = \frac{\sum_{t=1}^{T}(PM_{10}(t)-\mu_{PM_{10}})(P\hat{M}_{10}(t)-\mu_{P\hat{M}_{10}})}{\sqrt{\sum_{t=1}^{T}(PM_{10}(t)-\mu_{PM_{10}})^2}\cdot\sqrt{\sum_{t=1}^{T}(P\hat{M}_{10}(t)-\mu_{P\hat{M}_{10}})^2}}$$

where $P\hat{M}_{10}(t)$ and $PM_{10}(t)$ are, respectively, the t-th values of the model output and of the validation dataset, and $\mu_{P\hat{M}_{10}}$ and $\mu_{PM_{10}}$ are their mean values. From the huge set of performed tests, only the best-performing ones are presented in this context, in particular for the combination of multiple input.

3.3. Validation Results

3.3.1. Models with NO_2 as Input

This first class of models includes only NO_2 daily mean concentrations as input. This is due to the fact that PM_{10} and NO_2

concentrations are generated by several common emitting activities (i.e., road transport) and that the secondary inorganic fraction of PM_{10} is composed, in part, of nitrates, in particular ammonia nitrate, whose formation depends on the NO_2 concentration in

atmosphere. Table 1 highlights that the performances are quite good in terms of correlation, with values around 0.74, and acceptable in terms of root mean square error, with a normalised root mean standard deviation (allowing one to compare the root mean square error with respect to the overall variability of the output time series) around 0.1.

From these results, it is clear that an increase in the memory of the system does not lead to significant impacts on the performances and on the behavior of the model. The negligible increase in performances for the test with $n_{NO_2} = 4$ does not justify the increasing number of parameters. Table 2 shows the performances for the same configurations for the part of the time series where PM_{10} concentrations higher than 30 μg/m^3 have been measured. The table states that the model has strong difficulties in reproducing high concentrations, as highlighted by the strong decrease in statistical indexes.

3.3.2. Models with Meteorological Variables as Input

The second class of models considers only the meteorological variables as input. These tests allow an assessment of the relative "importance" between meteorology and NO_2 concentration for the computation of PM_{10} levels. Tables 3 and 4 show poor performances, with the limited exception of the cases with temperature T as input. Thus, the performances suggest that the meteorological conditions alone are not enough to estimate PM_{10} concentrations, and, so, they may be at best used to increase the performances in addition to the NO_2 concentrations.

Table 1. NO_2 input configuration performances.

	$x = \{NO_2\}$
	$n_{NO_2} = 1$
correlation	0.723
RMSE	10.896
NRMSD	0.1101
	$n_{NO_2} = 2$
correlation	0.722
RMSE	10.927
NRMSD	0.1104
	$n_{NO_2} = 3$
correlation	0.732
RMSE	10.764
NRMSD	0.1087
	$n_{NO_2} = 4$
correlation	0.733
RMSE	10.743
NRMSD	0.1085
	$n_{NO_2} = 5$
correlation	0.742
RMSE	10.575
NRMSD	0.1068

Table 2. NO_2 input configuration performances for $PM_{10} > 30\ \mu\mathrm{g/m^3}$.

		$x = \{NO_2\}$
		$n_{NO_2} = 1$
	correlation	0.39
	RMSE	11.531
	NRMSD	0.1696
		$n_{NO_2} = 2$
	correlation	0.39
	RMSE	11.481
	NRMSD	0.1688
		$n_{NO_2} = 3$
	correlation	0.41
	RMSE	11.390
	NRMSD	0.1675
		$n_{NO_2} = 4$
	correlation	0.40
	RMSE	11.412
	NRMSD	0.1678
		$n_{NO_2} = 5$
	correlation	0.40
	RMSE	11.444
	NRMSD	0.1683

Table 3. Meteorological input configuration performances.

	$x = \{WS\}$	$x = \{RF\}$	$x = \{RH\}$	$x = \{T\}$	$x = \{RH, T\}$
	$n_{WS} = 1$	$n_{RF} = 1$	$n_{RH} = 1$	$n_T = 1$	$n_{RH,T} = 1$
correlation	0.33	0.10	0.150	0.461	0.47
RMSE	16.733	16.622	16.380	14.807	14.736
NRMSD	0.1690	0.1679	0.1655	0.1496	0.1488
	$n_{WS} = 2$	$n_{RF} = 2$	$n_{RH} = 2$	$n_T = 2$	$n_{RH,T} = 2$
correlation	0.37	0.18	0.185	0.494	0.50
RMSE	16.723	16.486	16.132	14.551	14.469
NRMSD	0.1689	0.1665	0.1630	0.1470	0.1462
	$n_{WS} = 3$	$n_{RF} = 3$	$n_{RH} = 3$	$n_T = 3$	$n_{RH,T} = 3$
correlation	0.37	0.23	0.171	0.503	0.52
RMSE	16.899	16.283	16.176	14.467	14.385
NRMSD	0.1707	0.1644	0.1634	0.1461	0.1453
	$n_{WS} = 4$	$n_{RF} = 4$	$n_{RH} = 4$	$n_T = 4$	$n_{RH,T} = 4$
correlation	0.38	0.27	0.190	0.513	0.513
RMSE	16.912	15.9010	16.085	14.347	14.474
NRMSD	0.1708	0.1606	0.1625	0.1449	0.1462
	$n_{WS} = 5$	$n_{RF} = 5$	$n_{RH} = 5$	$n_T = 5$	$n_{RH,T} = 5$
correlation	0.339	0.30	0.183	0.514	0.52
RMSE	16.846	15.730	16.126	14.348	14.419
NRMSD	0.1701	0.1588	0.1629	0.1449	0.1456

Table 4. Meteorological input configuration performances for $PM_{10} > 30$ μg/m^3.

	$x = \{WS\}$	$x = \{RF\}$	$x = \{RH\}$	$x = \{T\}$	$x = \{RH, T\}$
	$n_{WS} = 1$	$n_{RF} = 1$	$n_{RH} = 1$	$n_T = 1$	$n_{UT} = 1$
correlation	0.19	0.079	0.15	0.35	0.32
RMSE	12.711	12.517	12.296	11.732	11.889
NRMSD	0.1869	0.1841	0.1808	0.1725	0.1748
	$n_{WS} = 2$	$n_{RF} = 2$	$n_{RH} = 2$	$n_T = 2$	$n_{UT} = 2$
correlation	0.196	0.153	0.18	0.40	0.43
RMSE	12.862	12.425	12.231	11.411	11.263
NRMSD	0.1891	0.1827	0.1799	0.1678	0.1656
	$n_{WS} = 3$	$n_{RF} = 3$	$n_{RH} = 3$	$n_T = 3$	$n_{UT} = 3$
correlation	0.185	0.212	0.20	0.43	0.40
RMSE	12.935	12.272	12.193	11.220	11.536
NRMSD	0.1902	0.1805	0.1793	0.1650	0.1696
	$n_{WS} = 4$	$n_{RF} = 4$	$n_{RH} = 4$	$n_T = 4$	$n_{UT} = 4$
correlation	0.24	0.210	0.17	0.44	0.42
RMSE	12.926	12.272	12.284	11.190	11.388
NRMSD	0.1901	0.1805	0.1806	0.1646	0.1675
	$n_{WS} = 5$	$n_{RF} = 5$	$n_{RH} = 5$	$n_T = 5$	$n_{UT} = 5$
correlation	0.192	0.220	0.20	0.44	0.45
RMSE	12.979	12.244	12.178	11.164	11.193
NRMSD	0.1909	0.1801	0.1791	0.1642	0.1646

3.3.3. Models with NO_2 and Meteorological Variables as Input

The last class of models considers both the meteorological variables and the NO_2 daily mean concentration as input in order to evaluate if the joint use of these information sources leads to an increase in the performances. Table 5 presents the results with NO_2 concentrations coupled to a meteorological variable at a certain time. The performances are in line with that of the models with only NO_2 as an input. Moreover, the combined use of more than one meteorological variable did not lead to a consistent increase in performance (Tables 6–8). The only slight improvement can be seen for high concentrations when the temperature is used as input (Tables 9–12), but also, in this case, the performances seem not to be good enough (correlation coefficient close to 0.52) in the preproduction of the peaks. These results suggest that, to reproduce mean PM_{10} levels in this domain, only the NO_2 concentrations should used, thus relying on cheaper sensors. Nevertheless, a bond in the performances exists, which did not allow the reconstruction of peak concentrations.

Table 5. NO_2 and one meteorological variable input configuration performance.

	$x = \{NO_2, WS\}$	$x = \{NO_2, RF\}$	$x = \{NO_2, RH\}$	$x = \{NO_2, T\}$
	$n_{NO_2,WS} = 1$	$n_{NO_2,RF} = 1$	$n_{NO_2,RH} = 1$	$n_{NO_2,T} = 1$
correlation	0.69	0.72	0.71	0.72
RMSE	11.285	10.987	10.969	10.808
NRMSD	0.1139	0.1109	0.1108	0.1092
	$n_{NO_2,WS} = 2$	$n_{NO_2,RF} = 2$	$n_{NO_2,RH} = 2$	$n_{NO_2,T} = 2$
correlation	0.70	0.73	0.72	0.73
RMSE	11.207	10.796	10.916	10.672
NRMSD	0.1132	0.1090	0.1103	0.1078
	$n_{NO_2,WS} = 3$	$n_{NO_2,RF} = 3$	$n_{NO_2,RH} = 3$	$n_{NO_2,T} = 3$
correlation	0.71	0.73	0.72	0.73
RMSE	11.042	10.735	10.890	10.669
NRMSD	0.1115	0.1084	0.1100	0.1078
	$n_{NO_2,WS} = 4$	$n_{NO_2,RF} = 4$	$n_{NO_2,RH} = 4$	$n_{NO_2,T} = 4$
correlation	0.71	0.731	0.72	0.74
RMSE	11.007	10.670	10.906	10.543
NRMSD	0.1111	0.1077	0.1102	0.1065
	$n_{NO_2,WS} = 5$	$n_{NO_2,RF} = 5$	$n_{NO_2,RH} = 5$	$n_{NO_2,T} = 5$
correlation	0.711	0.741	0.71	0.73
RMSE	10.950	10.500	10.954	10.716
NRMSD	0.1106	0.1060	0.1106	0.1082

Table 6. NO_2 and two meteorological variable input best configuration performances.

	$x = \{NO_2, WS, RF\}$	$x = \{NO_2, RH, T\}$	$x = \{NO_2, WS, T\}$	$x = \{NO_2, RF, RH\}$	$x = \{NO_2, RF, T\}$
	$n_{NO_2,WS,RF} = 1$	$n_{NO_2,RH,T} = 1$	$n_{NO_2,WS,T} = 1$	$n_{NO_2,RF,RH} = 1$	$n_{NO_2,RF,T} = 1$
correlation	0.70	0.72	0.70	0.72	0.73
RMSE	11.132	10.859	11.103	10.870	10.768
NRMSD	0.1124	0.1097	0.1122	0.1098	0.1088
	$n_{NO_2,WS,RF} = 2$	$n_{NO_2,RH,T} = 2$	$n_{NO_2,WS,T} = 2$	$n_{NO_2,RF,RH} = 2$	$n_{NO_2,RF,T} = 2$
correlation	0.72	0.73	0.69	0.725	0.73
RMSE	10.844	10.743	11.385	10.748	10.725
NRMSD	0.1095	0.1085	0.1150	0.1086	0.1083
	$n_{NO_2,WS,RF} = 3$	$n_{NO_2,RH,T} = 3$	$n_{NO_2,WS,T} = 3$	$n_{NO_2,RF,RH} = 3$	$n_{NO_2,RF,T} = 3$
correlation	0.711	0.73	0.72	0.72	0.73
RMSE	10.945	10.690	10.945	10.837	10.690
NRMSD	0.1105	0.1080	0.1106	0.1095	0.1080
	$n_{NO_2,WS,RF} = 4$	$n_{NO_2,RH,T} = 4$	$n_{NO_2,WS,T} = 4$	$n_{NO_2,RF,RH} = 4$	$n_{NO_2,RF,T} = 4$
correlation	0.72	0.72	0.73	0.71	0.73
RMSE	10.803	10.849	10.823	10.965	10.752
NRMSD	0.1091	0.1096	0.1093	0.1108	0.1086
	$n_{NO_2,WS,RF} = 5$	$n_{NO_2,RH,T} = 5$	$n_{NO_2,WS,T} = 5$	$n_{NO_2,RF,RH} = 5$	$n_{NO_2,RF,T} = 5$
correlation	0.731	0.72	0.72	0.70	0.71
RMSE	10.62	10.880	10.918	11.109	11.032
NRMSD	0.1072	0.1099	0.1103	0.1122	0.1114

Table 7. NO_2 and three meteorological variable input configuration performances.

	$x = \{NO_2, RF, RH, T\}$	$x = \{NO_2, WS, RF, RH\}$	$x = \{NO_2, WS, RF, T\}$
	$n_{NO_2,RF,RH,T} = 1$	$n_{NO_2,WS,RF,RH} = 1$	$n_{NO_2,WS,RF,T} = 1$
correlation	0.72	0.692	0.71
RMSE	10.841	11.241	10.992
NRMSD	0.1095	0.1135	0.1110
	$n_{NO_2,RF,RH,T} = 2$	$n_{NO_2,WS,RF,RH} = 2$	$n_{NO_2,WS,RF,T} = 2$
correlation	0.72	0.72	0.71
RMSE	10.826	10.868	11.064
NRMSD	0.1094	0.1098	0.1118
	$n_{NO_2,RF,RH,T} = 3$	$n_{NO_2,WS,RF,RH} = 3$	$n_{NO_2,WS,RF,T} = 3$
correlation	0.72	0.703	0.72
RMSE	10.869	11.084	10.916
NRMSD	0.1097	0.1119	0.1098
	$n_{NO_2,RF,RH,T} = 4$	$n_{NO_2,WS,RF,RH} = 4$	$n_{NO_2,WS,RF,T} = 4$
correlation	0.70	0.71	0.721
RMSE	11.098	10.032	11.163
NRMSD	0.1121	0.1114	0.1128
	$n_{NO_2,RF,RH,T} = 5$	$n_{NO_2,WS,RF,RH} = 5$	$n_{NO_2,WS,RF,T} = 5$
correlation	0.69	0.70	0.714
RMSE	11.324	11.207	11.071
NRMSD	0.1144	0.1132	0.1118

Table 8. NO_2 and four meteorological variable input configuration performances.

	$x = \{NO_2, WS, RF, RH, T\}$
	$n_{NO_2,WS,RF,RH,T} = 1$
correlation	0.70
RMSE	11.046
NRMSD	0.1116
	$n_{NO_2,WS,RF,U,T} = 2$
correlation	0.72
RMSE	10.926
NRMSD	0.1104
	$n_{NO_2,WS,RF,U,T} = 3$
correlation	0.73
RMSE	10.788
NRMSD	0.1090
	$n_{NO_2,WS,RF,U,T} = 4$
correlation	0.71
RMSE	11.632
NRMSD	0.1127
	$n_{NO_2,WS,RF,U,T} = 5$
correlation	0.71
RMSE	11.297
NRMSD	0.1141

Table 9. NO_2 one meteorological variable input configuration performance for $PM_{10} > 30$ μg/m^3.

	$x = \{NO_2, WS\}$	$x = \{NO_2, RF\}$	$x = \{NO_2, RH\}$	$x = \{NO_2, T\}$
	$n_{NO_2,WS} = 1$	$n_{NO_2,RF} = 1$	$n_{NO_2,RH} = 1$	$n_{NO_2,T} = 1$
correlation	0.381	0.39	0.392	0.39
RMSE	12.176	11.492	11.474	11.445
NRMSD	0.1791	0.1690	0.1687	0.1683
	$n_{NO_2,WS} = 2$	$n_{NO_2,RF} = 2$	$n_{NO_2,RH} = 2$	$n_{NO_2,T} = 2$
correlation	0.39	0.401	0.422	0.49
RMSE	11.970	11.456	11.280	10.817
NRMSD	0.1760	0.1685	0.1659	0.1591
	$n_{NO_2,WS} = 3$	$n_{NO_2,RF} = 3$	$n_{NO_2,RH} = 3$	$n_{NO_2,T} = 3$
correlation	0.382	0.413	0.406	0.50
RMSE	12.086	11.384	11.467	10.744
NRMSD	0.1777	0.1674	0.1686	0.1580
	$n_{NO_2,WS} = 4$	$n_{NO_2,RF} = 4$	$n_{NO_2,RH} = 4$	$n_{NO_2,T} = 4$
correlation	0.40	0.405	0.395	0.51
RMSE	11.938	11.438	11.553	10.671
NRMSD	0.1756	0.1682	0.1699	0.1569
	$n_{NO_2,WS} = 5$	$n_{NO_2,RF} = 5$	$n_{NO_2,RH} = 5$	$n_{NO_2,T} = 5$
correlation	0.40	0.425	0.402	0.52
RMSE	11.855	11.311	11.498	10.570
NRMSD	0.1743	0.1663	0.1691	0.1554

Table 10. NO_2 and two meteorological variable input best configuration performances for $PM_{10} > 30$ μg/m^3.

	$x = \{NO_2, WS, RF\}$	$x = \{NO_2, RH, T\}$	$x = \{NO_2, WS, T\}$	$x = \{NO_2, RF, RH\}$	$x = \{NO_2, RF, T\}$
	$n_{NO_2,WS,RF} = 1$	$n_{NO_2,U,T} = 1$	$n_{NO_2,WS,T} = 1$	$n_{NO_2,RF,U} = 1$	$n_{NO_2,RF,T} = 1$
correlation	0.394	0.42	0.45	0.394	0.44
RMSE	12.058	11.246	11.445	11.458	11.160
NRMSD	0.1773	0.1654	0.1683	0.1685	0.1641
	$n_{NO_2,WS} = 2$	$n_{NO_2,U,T} = 2$	$n_{NO_2,WS,T} = 2$	$n_{NO_2,RF,U} = 2$	$n_{NO_2,RF,T} = 2$
correlation	0.39	0.503	0.47	0.414	0.485
RMSE	11.970	10.750	11.367	11.364	10.832
NRMSD	0.1760	0.1581	0.1672	0.1671	0.1593
	$n_{NO_2,WS} = 3$	$n_{NO_2,U,T} = 3$	$n_{NO_2,WS,T} = 3$	$n_{NO_2,RF,U} = 3$	$n_{NO_2,RF,T} = 3$
correlation	0.382	0.49	0.47	0.40	0.49
RMSE	12.086	10.895	11.237	11.575	10.825
NRMSD	0.1777	0.1602	0.1652	0.1702	0.1592
	$n_{NO_2,WS} = 4$	$n_{NO_2,U,T} = 4$	$n_{NO_2,WS,T} = 4$	$n_{NO_2,RF,U} = 4$	$n_{NO_2,RF,T} = 4$
correlation	0.40	0.50	0.51	0.40	0.51
RMSE	11.938	10.919	10.839	11.658	10.667
NRMSD	0.1756	0.1606	0.1594	0.1714	0.1569
	$n_{NO_2,WS} = 5$	$n_{NO_2,U,T} = 5$	$n_{NO_2,WS,T} = 5$	$n_{NO_2,RF,U} = 5$	$n_{NO_2,RF,T} = 5$
correlation	0.40	0.514	0.52	0.40	0.49
RMSE	11.855	10.707	10.703	11.632	10.856
NRMSD	0.1743	0.1575	0.1574	0.1711	0.1597

Table 11. NO_2 and three meteorological variable input configuration performances for $PM_{10} > 30$ μg/m^3.

	$x = \{NO_2, RF, RH, T\}$	$x = \{NO_2, WS, RF, RH\}$	$x = \{NO_2, WS, RF, T\}$
	$n_{NO_2,RF,RH,T} = 1$	$n_{NO_2,WS,RF,RH} = 1$	$n_{NO_2,WS,RF,T} = 1$
correlation	0.430	0.390	0.436
RMSE	11.201	11.660	11.444
NRMSD	0.1647	0.1715	0.1683
	$n_{NO_2,RF,RH,T} = 2$	$n_{NO_2,WS,RF,RH} = 2$	$n_{NO_2,WS,RF,T} = 2$
correlation	0.503	0.407	0.462
RMSE	10.720	11.863	11.275
NRMSD	0.1576	0.1744	0.1658
	$n_{NO_2,RF,RH,T} = 3$	$n_{NO_2,WS,RF,RH} = 3$	$n_{NO_2,WS,RF,T} = 3$
correlation	0.49	0.381	0.478
RMSE	10.945	12.103	11.063
NRMSD	0.1610	0.1780	0.1627
	$n_{NO_2,RF,RH,T} = 4$	$n_{NO_2,WS,RF,RH} = 4$	$n_{NO_2,WS,RF,T} = 4$
correlation	0.503	0.396	0.498
RMSE	10.892	12.077	10.846
NRMSD	0.1602	0.1776	0.1595
	$n_{NO_2,RF,RH,T} = 5$	$n_{NO_2,WS,RF,RH} = 5$	$n_{NO_2,WS,RF,T} = 5$
correlation	0.511	0.399	0.518
RMSE	10.756	11.953	10.678
NRMSD	0.1582	0.1758	0.1570

Table 12. NO_2 and four meteorological variable input configuration performances for $PM_{10} > 30$ μg/m^3.

	$x = \{NO_2, WS, RF, RH, T\}$
	$n_{NO_2,WS,RF,RH,T} = 1$
correlation	0.425
RMSE	11.484
NRMSD	0.1689
	$n_{NO_2,WS,RF,RH,T} = 2$
correlation	0.494
RMSE	11.067
NRMSD	0.1627
	$n_{NO_2,WS,RF,RH,T} = 3$
correlation	0.484
RMSE	11.191
NRMSD	0.1646
	$n_{NO_2,WS,RF,RH,T} = 4$
correlation	0.492
RMSE	11.107
NRMSD	0.1633
	$n_{NO_2,WS,RF,RH,T} = 5$
correlation	0.522
RMSE	10.875
NRMSD	0.1599

3.4. Comparison to State-of-the-Art Models

In this section, the comparison of the wavenet approach used in this work with two different state-of-the-art models is presented. The two models are a (1) K-nearest neighbors (KNN) and an (2) artificial neural network-based model, which are often used in this context to capture the dynamic of the PM_{10} [29]. The comparison (Table 13) shows how the performances of the best-identified wavenet are strongly better than that of the KNN model and very similar (slightly better for high orders) to that of the ANN ones. Moreover, it has to be stressed how the best model for the wavenet approach ensures these performances with limited complexity and with a limited number of variables (only NO_2 concentration) with respect to the other approaches.

Figures 5–7 present the time series plots for the best configuration of wavenet, artificial neural network and KNN models, respectively. As expected, the behaviour of the wavenet and ANN models is very similar, with the first models showing slightly better performances for the low value close to the sample n. 800. In general, the KNN model reproduces higher value but, as also stated by the lower values of correlation coefficient, the time series rarely follows the value and the gradient of the measured values.

Table 13. Best configuration performances.

|	WT ($x = \{NO_2\}$)	ANN ($x = \{NO_2, RF\}$)	KNN ($x = \{NO_2, WS, RF, U\}$)
	$n_{NO_2} = 1$	$n_{NO_2,RF} = 1$	$n_{NO_2,WS,RF,U} = 1$
correlation	0.723	0.72	0.523
RMSE	10.896	10.940	14.860
NRMSD	0.1101	0.1105	0.1501
	$n_{NO_2} = 2$	$n_{NO_2,RF} = 2$	$n_{NO_2,WS,RF,U} = 2$
correlation	0.722	0.732	0.581
RMSE	10.927	10.651	14.082
NRMSD	0.1104	0.1076	0.1422
	$n_{NO_2} = 3$	$n_{NO_2,RF} = 3$	$n_{NO_2,WS,RF,U} = 3$
correlation	0.732	0.72	0.59
RMSE	10.764	10.877	13.904
NRMSD	0.1087	0.1099	0.1404
	$n_{NO_2} = 4$	$n_{NO_2,RF} = 4$	$n_{NO_2,WS,RF,U} = 4$
correlation	0.733	0.713	0.61
RMSE	10.743	10.909	13.628
NRMSD	0.1085	0.1102	0.1377
	$n_{NO_2} = 5$	$n_{NO_2,RF} = 5$	$n_{NO_2,WS,RF,U} = 5$
correlation	0.742	0.73	0.634
RMSE	10.575	10.675	13.435
NRMSD	0.1068	0.1078	0.1357
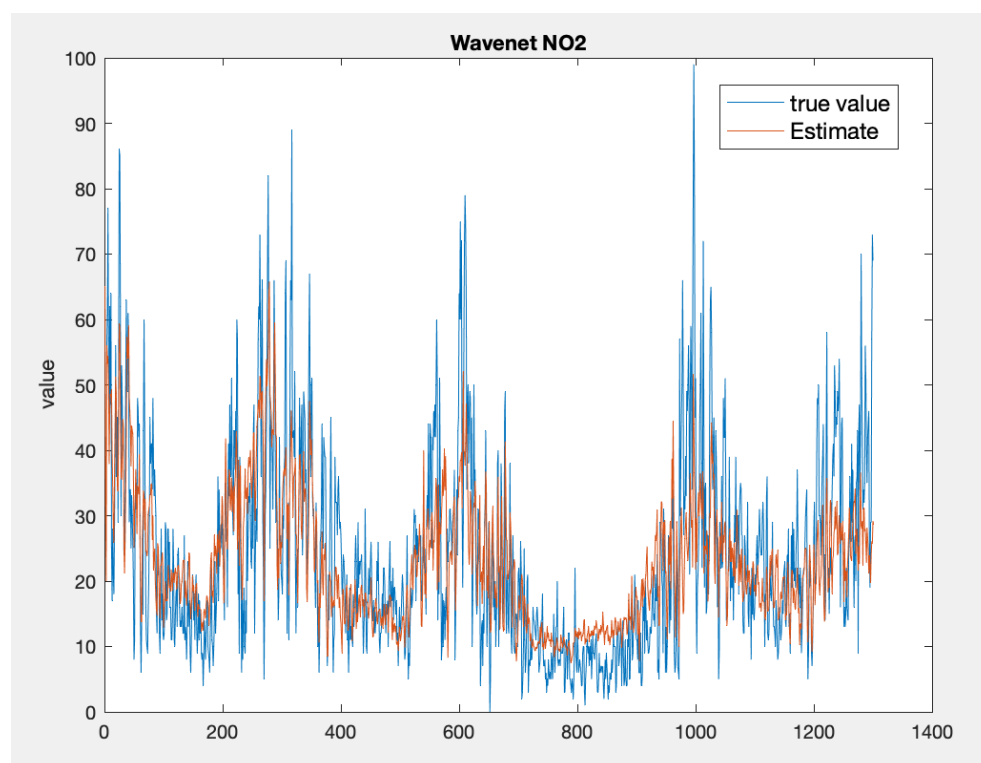
Figure 5. Time series comparison between the measured values (blue) and the best wavenet model output ($n_{NO_2} = 5$, red).
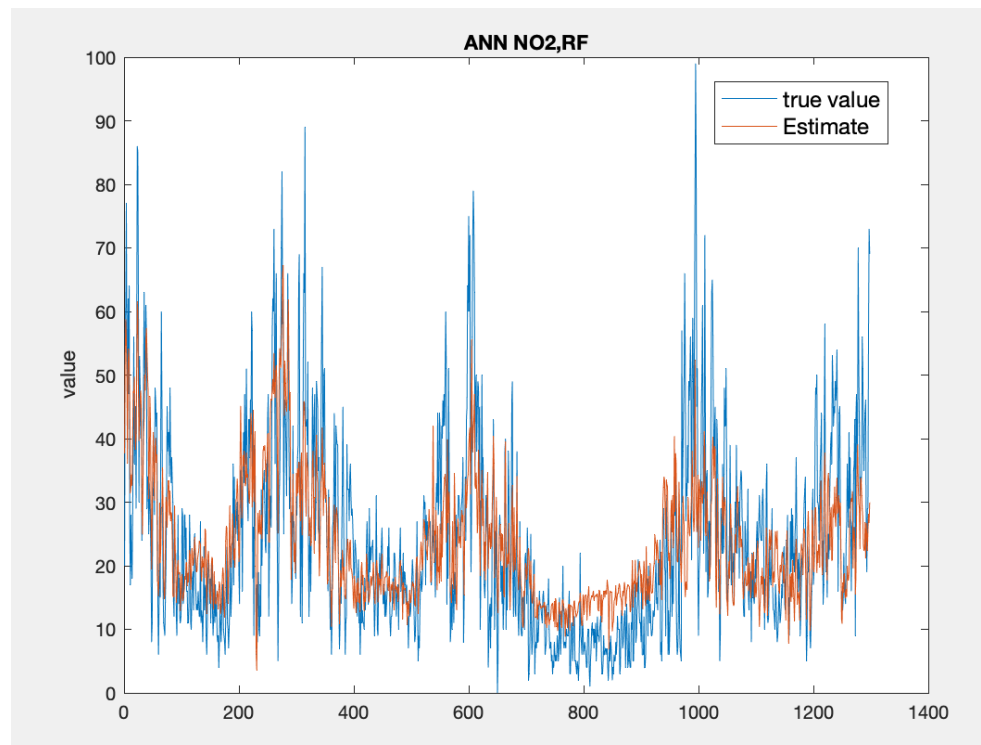

Figure 6. Time series comparison between the measured values (blue) and the best neural network model output ($n_{NO_2,RF} = 2$, red).
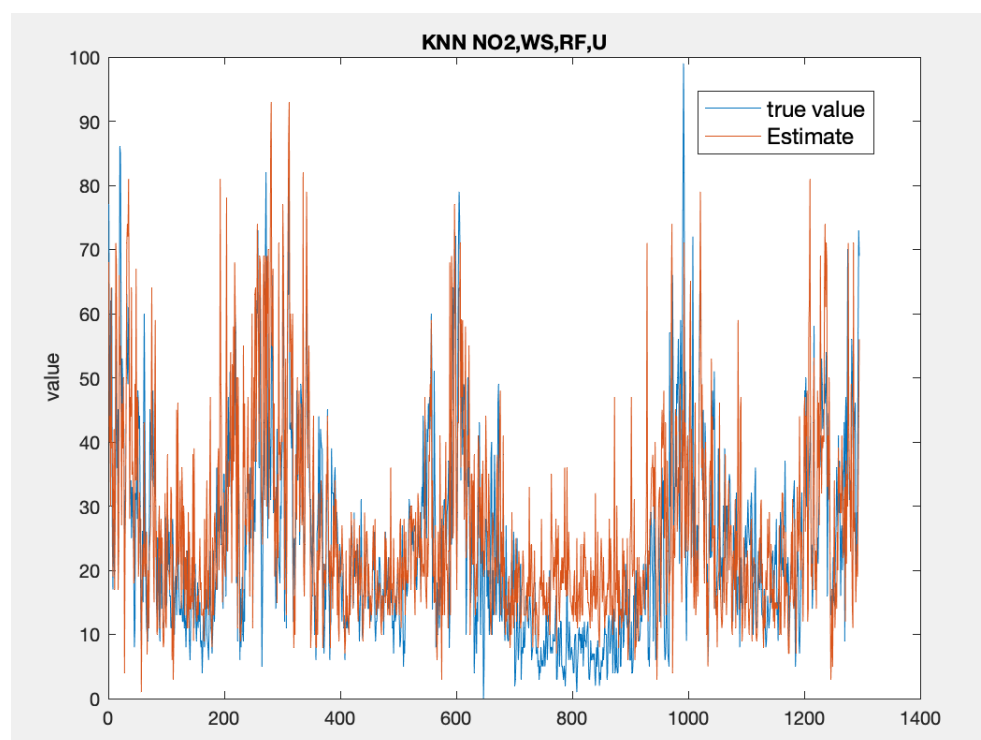
Figure 7. Time series comparison between the measured values (blue) and the best KNN model output ($n_{NO_2,RF} = 5$, red).

4. Conclusions

In this work, a data-driven, wavenet-based virtual sensor for PM_{10} daily mean concentration is presented and evaluated. Different model configurations have been tested and evaluated. The methodology has been applied to data measured by the Lombardy regional monitoring network. The results show good agreement between the output of the virtual sensor and the measured data used for validation when the daily mean NO_2 concentration is used as input—in particular, around the mean concentration values. Therefore, the models fail to reproduce the peak concentrations, and this behaviour will not change even if other inputs, such as meteorological data, are used. Nevertheless, the performances show that this approach can be used to produce supporting information to integrate the regional monitoring network that can be made available through app/web services due to a relatively fast computation.

Author Contributions: Conceptualization, C.C., E.T., R.Z.; software, C.C., R.Z.; validation, C.C., E.D.A., M.V.; funding acquisition, M.V. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Landrigan, P.; Fuller, R.; Acosta, N.; Adeyi, O.; Arnold, R.; Basu, N.; Baldé, A.; Bertollini, R.; Bose-O'Reilly, S.; Boufford, J.; et al. The Lancet Commission on pollution and health. *Lancet* **2018**, *391*, 462–512. [CrossRef]
2. Pope, C., III; Dockery, D.; Spengler, J.; Raizenne, M. Respiratory health and PM$_{10}$ pollution: A daily time series analysis. *Am. Rev. Respir. Dis.* **1991**, *144*, 668–674. [CrossRef]
3. Pope, C., III; Dockery, D. Acute health effects of PM$_{10}$ pollution on symptomatic and asymptomatic children. *Am. Rev. Respir. Dis.* **1992**, *145*, 1123–1128. [CrossRef] [PubMed]
4. Turrini, E.; Carnevale, C.; Finzi, G.; Volta, M. A non-linear optimization programming model for air quality planning including co-benefits for GHG emissions. *Sci. Total Environ.* **2018**, *621*, 980–989. [CrossRef]

5.   Carnevale, C.; Ferrari, F.; Guariso, G.; Maffeis, G.; Turrini, E.; Volta, M. Assessing the Economic and Environmental Sustainability of a Regional Air Quality Plan. *Sustainability* **2018**, *10*, 3568. [CrossRef]

6.   Gimez Vilchez, J.; Julea, A.; Peduzzi, E.; Pisoni, E.; Krause, J.; Siskos, P.; Thiel, C. Modelling the impacts of EU countries electric car deployment plans on atmospheric emissions and concentrations. *Eur. Transp. Res. Rev.* **2019**, *11*, 40. [CrossRef]

7.   Relvas, H.; Miranda, A.; Carnevale, C.; Maffeis, G.; Turrini, E.; Volta, M. Optimal air quality policies and health: A multi-objective nonlinear approach. *Environ. Sci. Pollut. Res.* **2017**, *24*, 13687–13699. [CrossRef]

8.   Marques, G.; Pires, I.M.; Miranda, N.; Pitarma, R. Air Quality Monitoring Using Assistive Robots for Ambient Assisted Living and Enhanced Living Environments through Internet of Things. *Electronics* **2019**, *8*, 1375. [CrossRef]

9.   Arroyo, P.; Lozano, J.; Suárez, J. Evolution of Wireless Sensor Network for Air Quality Measurements. *Electronics* **2018**, *7*, 342. [CrossRef]

10.   Carnevale, C.; Finzi, G.; Pederzoli, A.; Pisoni, E.; Thunis, P.; Turrini, E.; Volta, M. A methodology for the evaluation of re-analyzed PM10 concentration fields: A case study over the PO Valley. *Air Qual. Atmos. Health* **2015**, *8*, 533–544. [CrossRef]

11.   Candiani, G.; Carnevale, C.; Finzi, G.; Pisoni, E.; Volta, M. A comparison of reanalysis techniques: Applying optimal interpolation and Ensemble Kalman Filtering to improve air quality monitoring at mesoscale. *Sci. Total Environ.* **2013**, *458–460*, 7–14. [CrossRef]

12.   Carnevale, C.; Finzi, G.; Pederzoli, A.; Turrini, E.; Volta, M.; Ferrari, F.; Gianfreda, R.; Maffeis, G. Impact of pollutant emission reductions on summertime aerosol feedbacks: A case study over the PO valley. *Atmos. Environ.* **2015**, *122*, 41–57. [CrossRef]

13.   Winkel, A.; Llorens Rubio, J.; Huis in't Veld, J.W.; Vonk, J.; Ogink, N.W. Equivalence testing of filter-based, beta-attenuation, TEOM, and light-scattering devices for measurement of PM10 concentration in animal houses. *J. Aerosol. Sci.* **2015**, *80*, 11–26. [CrossRef]

14.   Costa, A.; Guarino, M. Definition of yearly emission factor of dust and greenhouse gases through continuous measurements in swine husbandry. *Atmos. Environ.* **2009**, *43*, 1548–1556. [CrossRef]

15.   Pathak, D.; Halale, V.P. An Introductory Approach to Virtual Sensors and Its Modelling Techniques. *Sci. Eng. Res.* **2016**, *7*, 461–464.

16.   Sun, S.; He, Y.; Zhou, S.D.; Yue, Z.J. A Data-Driven Response Virtual Sensor Technique with Partial Vibration Measurements Using Convolutional Neural Network. *Sensors* **2017**, *17*, 2888. [CrossRef]

17.   Kuo, S.; Zhou, M. Virtual sensing techniques and their applications. In Proceedings of the International Conference on Networking, Sensing and Control, Okayama, Japan, 26–29 March 2009; pp. 31–39

18.   Rai, A.C.; Kumar, P.; Pilla, F.; Skouloudis, A.N.; Sabatino, S.D.; Ratti, C.; Yasar, A.; Rickerby, D. End-user perspective of low-cost sensors for outdoor air pollution monitoring. *Sci. Total Environ.* **2017**, *607–608*, 691–705. [CrossRef]

19.   Saljooghi, S.; Hezarkarkhani, A. Comparison of WAVENET and ANN for predicting the porosity obtained from well log data. *J. Petrol. Sci. Eng.* **2014**, *123*, 172–182. [CrossRef]

20.   van den Oord, A.; Dieleman, S.; Zen, H.; Simonyan, K.; Vinyals, O.; Graves, A.; Kalchbrenner, N.; Senior, A.; Kavukcuoglu, K. WaveNet: A Generative Model for Raw Audio. *arXiv* **2016**, arXiv:1609.03499.

21.   Saljooghi, S.; Hezarkarkhani, A. A new approach to improve permeability prediction of petroleum reservois using neural network adaptive wavelet (Wavenet). *J. Petrol. Sci. Eng.* **2015**, *123*, 851–861. [CrossRef]

22.   Carnevale, C.; Angelis, E.; Finzi, G.; Turrini, E.; Volta, M. Application of data fusion techniques to improve air quality forecast: A case study in the Northern Italy. *Atmosphere* **2020**, *11*, 244. [CrossRef]

23.   Carnevale, C.; De Angelis, E.; Tagliani, F.; Turrini, E.; Volta, M. A short-term air quality control for PM_{10} levels. *Electronics* **2020**, *9*, 1409. [CrossRef]

24.   Sujatha, P. *Vibration and Acoustics: Measurement and Signal Analysis*; McGraw-Hill Education: Gautam Buddha Nagar, India, 2010.

25.   Cybenko, G. Approximation by Superposition of a Sigmoidal Function. *Math. Control Signals Syst.* **1989**, *2*, 303–314. [CrossRef]

26.   Mallat, S. A Theory for Multiresolution Signal Decomposition The Wavelet Representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 674–693. [CrossRef]

27.   Zakeri, V.; Naghavi, V.; Safavi, A. Developing real-time wave-net models for non-linear time-varying experimental processes. *Comput. Chem. Eng.* **2009**, *33*, 1379–1385. [CrossRef]

28.   Postalciouglu, S.; Erkan, K.; Bolat, E.D. Comparison of Wavenet and Neuralnet for System Modeling. In *Knowledge-Based Intelligent Information and Engineering Systems*; Khosla, R., Howlett, R.J., Jain, L.C., Eds.; Springer: Berlin/Heidelberg, Germany, 2005; pp. 100–107.

29.   Carnevale, C.; Finzi, G.; Guariso, G.; Pisoni, E.; Volta, M. Surrogate models to compute optimal air quality planning policies at a regional scale. *Environ. Model. Softw.* **2012**, *34*, 44–50. [CrossRef]