Recurrent Air Quality Predictor Based on Meteorology- and Pollution-Related Factors

Ke Gu, Junfei Qiao, *Member, IEEE*, and Weisi Lin, *Fellow, IEEE*

***Abstract*—Air quality is currently arousing drastically increasing attention from the governments and populace all over the world. In this paper, we propose a heuristic recurrent air quality predictor (RAQP) to infer air quality. The RAQP exploits some key meteorology- and pollution-related variables to infer air pollutant concentrations (APCs), e.g. the fine particulate matter (PM2.5). It is natural that the meteorological factors and APCs at the current time have strong influences on air quality the next adjacent moment, that is to say, there exist high correlations between them. With this consideration, applying simple machine learners to the current meteorology- and pollution-related factors can reliably predict the air quality indices at a time later. However, owing to the nonlinear and chaotic reasons, the above correlations decline with the time interval enlarged. In such cases, it fails to forecast the air quality after several hours by only using simple machine learners and the current measurements of meteorology- and pollution-related variables. To solve the problem, our RAQP method recurrently applies the 1-h prediction model, which learns the current records of meteorology- and pollution-related factors to predict the air quality 1 h later, to then estimate the air quality after several hours. Via extensive experiments, results confirm that the RAQP predictor is superior to the relevant state-of-the-art techniques and nonrecurrent methods when applied to air quality prediction.**

***Index Terms*—Air pollutant concentrations (APCs), Air quality prediction, meteorological factors (MFs), recurrent, regression.**

I. Introduction

RECENT decades have witnessed the quick urbanization and industrialization, inevitably along with remarkably and constantly rising air, water, and food pollution, in many regions, particularly in China. As compared with the problem caused by polluted water and food that the folk may solve by introducing more forceful tactics, such as adding larger doses of the agent for sewage disposal, humans seem fairly helpless when facing the ubiquitous polluted air. The "Great Smog of London," which killed thousands of persons in only a few days, is still fresh in our memories. Also, it is worthy to emphasize that it is generally accompanied with the high-density crowded population in those rapidly developed areas. Therefore, how to validly prevent such a crowd of people from the danger of air pollution and guarantee them in good health is highly concerned by the governments at the present time, and evidently, this concern will be continually amplified since treating air pollution is of extreme difficulty, and it cannot be totally resolved in a short time. With the above consideration, an efficient and effective predictor to forecast the air quality during the next several hours is eagerly desired, which will substantially facilitate the decision making of the government, e.g., traffic restriction, toward reducing the exhaust emissions discharged to the atmosphere.

One critical way to give rise to poor air quality is due to the anthropogenic-caused particulate and gaseous emissions, which typically include motor vehicles, industrial processes, coal, oil and natural gas combustion, etc. [1]. The commonly seen harmful air pollutants are composed of NO_2, O_3, CO, and so forth. Beyond a certain concentration, the former two pollutants are easy to bring about respiratory inflammation, while the third one might even damage blood and nervous system and therefore cause the body death. Apart from those above air pollutants, a growing number of attention has been concentrated on the fine particulate matter (PM2.5), which is a complicated air pollutant mixed with particles beneath the aerodynamic diameters of 2.5 μm. Another similar pollutant, called the inhalable particles (PM10), is composed of particles with the aerodynamic diameters of 10 μm or smaller [2]. In contrast to PM10, the governments and folks focus more on PM2.5 since it is easier to invade and lodge deeply into the lungs, and this undoubtedly leads the increased morbidity and mortality to the public under the condition of chronic exposure to high-concentration PM2.5.

We take PM2.5, one of the most concerned air pollutants, for example. Which meteorological factors (MFs) impact the PM2.5 concentration and how to influence them remain under exploration. To our best knowledge, it can be generally acknowledged that the variation of PM2.5 concentration is jointly determined by a series of MFs. For instance, some critical studies unveiled that the aerosol optical thickness (AOT) is closely correlated with the PM2.5 concentration [3]–[5], even nicely fitted with a simple linear model under some special cases [6]. Regardless of from research works or experiences, an apparent influence of high-concentration PM2.5 is to cause a severe

Manuscript received October 15, 2017; revised December 12, 2017; accepted December 27, 2017. Date of publication January 15, 2018; date of current version September 4, 2018. This work was supported in part by the National Science Foundation of China under Grant 61703009. Paper no. TII-17-2403.R1. *(Corresponding author: Ke Gu.)*

K. Gu and J.-F. Qiao are with the Beijing Key Laboratory of Computational Intelligence and Intelligent System, Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China (e-mail: guke.doctor@gmail.com; junfeiq@bjut.edu.cn).

W. Lin is with the School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798 (e-mail: wslin@ntu.edu.sg).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TII.2018.2793950

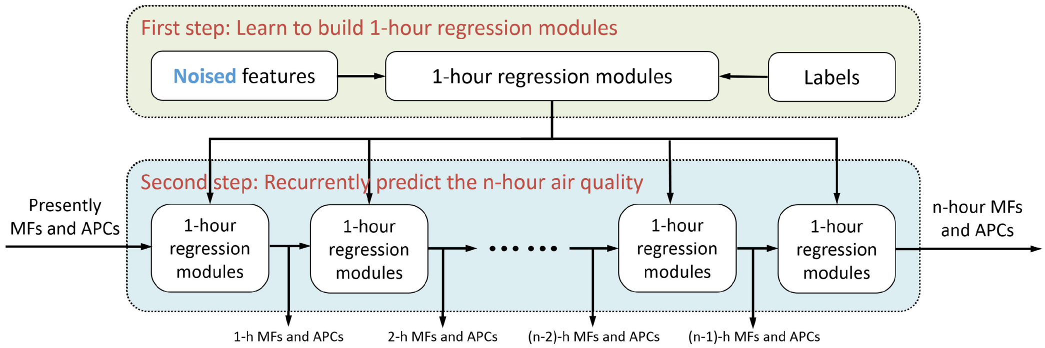


Fig. 1. Proposed recurrent model for hourly predictions of air pollutants.

visibility decrease [7]–[9]. How typical meteorological parameters, e.g., wind speed, relative humidity, temperature, and so forth, affect the PM2.5 concentration was investigated as well, and they were found to have fairly high positive or negative correlations [10]–[14]. Besides, some recent studies have shown that the majority of air pollutants are strongly correlated with each other [15].

To unearth and model the complex nonlinear relationship of real-world processes, such as estimating the air quality, was found to be a tough mission [16], [17]. One classical solution relies on the process-based approaches, which are devoted to modeling atmospheric and chemical processes encompassed in the air pollutant production with some chemical transport models (CTMs) [18]. Nonetheless, in this type of approaches, there exists one critical problem that the CTMs are generally too complicated to be modeled even though a huge massive of detailed information, e.g., meteorology knowledge and gas emissions inventories, can be available, let alone without the information that is hard to be accessed in real applications. To address this problem, another solution was proposed based on data-driven statistical approaches, which resort to ground monitoring measurements or satellite readings and therefore modeling the air pollution process is not necessarily required [19]. Under special conditions that a near-linear relationship exists between MFs and air quality prediction, linear regression models can be adopted in those data-driven approaches to contribute a good result [20]. However, most real-world processes are nonlinear, and thus advanced nonlinear regression models, e.g. neural network, are needed to establish a mapping from input meteorological parameters to the output air quality estimation. In some works [21]–[23], nonlinear regression models were demonstrated to be better applied to air pollution modeling. In [24], the authors further incorporated the linear regression model, neural network, and persistence model together for accurately forecasting the daily mean of PM2.5 concentrations on the US−Mexico border. In recent works [25]–[27], several advanced machine learning tools have been successfully applied to air quality prediction. We will illustrate these models in the section of experimental results since they are included for performance comparison.

In this paper, we attempt to use the MFs and air pollutant concentrations (APCs) acquired at the current moment to forecast the hourly estimations of APCs numerous hours later. To solve the aforementioned problem, we propose a general-purpose framework, which is capable of amending the performance of simple learning-based models or existing air quality prediction models to a sizable margin. For illustration, consider deploying the popular support vector regressor (SVR) [28] to learn the presently MFs and APCs.[1] Via experiments, it can be observed that, in the short term, the prediction performance is fairly good because the MFs and APCs are closely correlated with APCs to be predicted, whereas the correlation drops promptly and drastically as the time interval increases. Due to the weak correlation, directly using the SVR to learn the predictions of APCs is unreliable. Instead, we introduce a recurrent framework to address the above problem and furthermore combine the framework and SVR to develop a recurrent air quality predictor (RAQP). Particularly, when the presently MFs and APCs are used to predict the APCs after n hours ($n > 1$), we firstly predict the MFs and APCs after 1 h followed by utilizing the 1-h prediction model and 1-h predicted outputs of MFs and APCs as the input to infer the MFs and APCs after 2 h. Recurrently implementing the aforesaid process until the n-h MFs and APCs are estimated. Note that there must exist errors between the intermediate outputs and the unknown truth inputs because the 1-h prediction model cannot be 100% accurate, which leads to the error accumulation and makes the 1-h prediction model unable to work. For this, we might as well suppose the error is small enough to be ignored. In fact, a better solution will be provided in the next paragraph. Fig. 1 illustrates the proposed framework: 1) learning the regression modules between the MFs and APCs at the present time and those after 1 h; 2) recurrently using the 1-h regression modules trained above to estimate the MFs and APCs after numerous hours.

Compared with the previous works, this paper has the two main contributions. *First*, to the best of our knowledge, this paper is the first one that applies the recurrent strategy to air quality prediction. The proposed recurrent-based RAQP model is not only a predictor, but also provides a general-purpose framework that is applicable to raising the performance of simple learning-based models or existing air quality predictors. *Second*, we are the first using the noised features (as labeled in Fig. 1) when inferring the air quality. To specify, we add 100 randomly generated noise sets to the training feature set and therefore make the

[1]Other learning models or air quality predictors will be used to check the effectiveness of the proposed framework. Please see Section III.

training samples 100 times larger. By manually injecting noise, the generalization of the regression modules will be enhanced since the number of training samples is largely increased and meanwhile in real applications the noises must be included in the measurements of MFs and APCs obtained from instruments. Furthermore, it deserves to stress that the trained 1-h prediction model is immune to the errors (even totally when the 1-h prediction model is near to perfect or the noise variance is large), so it can be directly used to learn the intermediate outputs without leading to the serious error accumulation.

The structure of the remainder of this paper is outlined as follows. Section II introduces the proposed RAQP predictor. In Section III, the performance of the recurrent-based RAQP model is compared with the state-of-the-art competitors and nonrecurrent methods. We conclude the paper in Section IV.

II. Air Quality Prediction

Quality diagnosis and monitoring have long played critical roles in typical industrial applications, which include power systems [29], networks [30], [31], video technologies [32]–[35], electric vehicles [36], etc, and the relevant subsequent quality controlling and improvement have also attracted an extensive scope of attention from the industrial society [37], [38]. On the one hand, with the high-speed development of scientific technology, particularly during the recent decades when great achievements have been made in various kinds of application scenarios, an increasing number of advanced technologies are being applied to human health and longevity; on the other hand, risingly wider range of pollution of air, water, and food is accompanied with the rapid urbanization and industrialization, which is also highly concerned by the governments and folks at the present time. Thus, an efficient and effective air quality prediction model will be an urgent and crucial task in the next few decades or longer.

A. Direct Prediction

The support vector machine (SVM) was first explored by the AT&T Bell Laboratories for the purpose of classification, and later on was further extended to SVR for the regression problems [39]. Supposing a training date set $S = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_m, y_m)\}$, where $\mathbf{x}_i = [x_i^1, x_i^2, \ldots, x_i^n]^T \in R^n$ is the ith vector of feature inputs and $y_i \in R$ is the ith real target output. We can present the general form of the SVR using a hyperplane function:

$$h(\mathbf{x}_i) = \langle \boldsymbol{\alpha}, \mathcal{X}(\mathbf{x}_i) \rangle + \beta \tag{1}$$

where $\langle \cdot, \cdot \rangle$ indicates the inner product; $\mathcal{X}(\cdot)$ denotes a nonlinear function applied to the feature space; $\boldsymbol{\alpha}$ and β stand for the parameters to be determined based on the given feature inputs and target output. Minimum flatness of the function h can be transferred to minimizing the norm of $\|\boldsymbol{\alpha}\|_2^2$. In real applications, the slack variables $\boldsymbol{\zeta}$ and $\boldsymbol{\zeta}^*$ are supplemented to account for some margin of the errors. Via the analysis above, we can derive

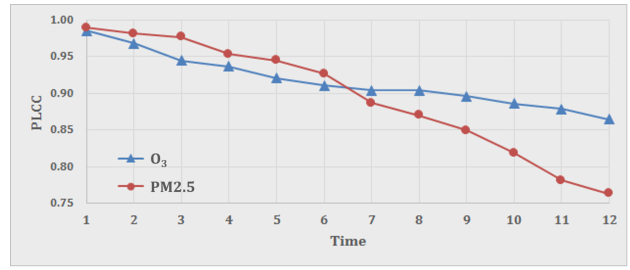


Fig. 2. Prediction accuracy of PM2.5 and O_3 from T_1 to T_{12}.

the convex optimization problem:

$$\begin{aligned} &\text{minimize} \quad \frac{1}{2}\|\boldsymbol{\alpha}\|_2^2 + \lambda \sum_{i=1}^{m} (\zeta_i + \zeta_i^*) \\ &\text{subject to} \begin{cases} \langle \boldsymbol{\alpha}, \mathcal{X}(\mathbf{x}_i) \rangle + \beta - y_i \le e + \zeta_i \\ y_i - \langle \boldsymbol{\alpha}, \mathcal{X}(\mathbf{x}_i) \rangle - \beta \le e + \zeta_i^* \\ \zeta_i, \zeta_i^* \ge 0, \; i = 1, 2, \ldots, m \end{cases} \end{aligned} \tag{2}$$

where e is the error tolerance range of the approximating function; λ represents a regularization parameter no less than zero, used for regulating the flatness of the function h and tolerance limits of the error beyond e. The constraints above guarantee that the majority of the data $\mathbf{x}_i$ are located in the tube $|y_i - \langle \boldsymbol{\alpha}, \mathcal{X}(\mathbf{x}_i) \rangle - \beta| \le e$. Otherwise, if $\mathbf{x}_i$ exceeds the tube, an error $\boldsymbol{\zeta}$ or $\boldsymbol{\zeta}^*$ will be yielded and minimized in the objective function. In general, we minimize the regularization term $\frac{1}{2}\|\boldsymbol{\alpha}\|_2^2$ and the error term $\lambda \sum_{i=1}^{m} (\zeta_i + \zeta_i^*)$ to address the underfitting and overfitting issues. We define the kernel function $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle$, which is employed for mapping the data $\mathbf{x}$ to a higher dimensional space. Here, the commonly used radial basis function (RBF) kernel, denoted as $\mathcal{K}(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2)$, is applied in our work. Using the training samples, we expect to find the parameters λ, e, and γ and thus determine the regression model.

We roughly examine the performance of SVR-based direct prediction model, which uses the records of MFs and APCs at T_0 to forecast the air quality indices from T_1 to T_{12}. T_0 means an initial time and T_n ($n \geqslant 1$) means the nth hour after T_0. We randomly divide the entire air quality prediction dataset into two classes. One class contains 80% data for training and the other contains the rest 20% data for testing. We repeat the above process 100 times and use the popular Pearson linear correlation coefficient (PLCC) to measure the average prediction accuracy, as given in Fig. 2. Red and blue dots, respectively, correspond to PM2.5 and O_3. Larger PLCC values indicate better prediction accuracy. More illustrations regarding the testing dataset and how to compute the PLCC index will be detailedly described in the next section. As can be observed from Fig. 2, regardless of PM2.5 and O_3, the PLCC value dramatically decreases as the time interval grows. This indicates that the direct strategy is good at short-term predictions of air quality, whereas it fails on mid- and long-term predictions, which is very possibly because of the weak correlations between the currently records of MFs and APCs and the true values of APCs after a long term.

B. Recurrent Prediction

Aiming to address the problem mentioned above, in this paper, we propose a heuristic solution with a recurrent strategy and the associated RAQP predictor. Note that, when the time interval is small, i.e., as for short-term predictions, the prediction performance can reach to a high level because there exist strong correlations between the measurements of the input MFs and APCs and the true values of APCs to be predicted. Therefore, it is natural to associate a recurrent strategy, to specify, which recurrently adopts the short-term prediction model to infer the mid- and long-term air quality indices. The aforesaid strategy is not a naive and groundless idea but inspired from some classical industrial technologies that have been widely used in numerous other application fields. We take the compression technologies of, e.g., power, circuit, acoustic, video, and heart rate signals for example. We define that the values of a signal vector to be compressed are, respectively, θ_i and θ_j at the T_i and T_j moments, where $j > i$, $P_j(\theta_i)$ is the predicted value of θ_j based on θ_i, and the difference between $P_j(\theta_i)$ and θ_j is $\epsilon_{i\to j}$. Supposing that θ_0 is known, the direct strategy for compressing a signal is to estimate and save the error difference between $P_i(\theta_0)$ and the following θ_i, namely $\epsilon_{0\to i}$, where $i > 0$. By comparison, the recurrent-based compression technology, which predicts and saves the difference between $P_{i+1}(\theta_i)$ and θ_{i+1}, i.e., $\epsilon_{i\to i+1}$, where i starts from 0 until the whole signal ends, drastically outperforms the direct strategy, because a signal value and its predicted version based on the neighboring value are closely correlated with each other and therefore $\sum_i \epsilon_{i\to i+1}$ is much less than $\sum_i \epsilon_{0\to i}$. The estimation of error difference in the compression methods is similar to and can be extended to the air quality estimation. Due to the high correlation during the neighboring moments, the prediction models might reach to the perfect 100% performance, i.e., $\epsilon_{i\to i+1} = 0$ for $\forall i$, which means that we can accurately predict the next-moment values. Extended to the air quality estimation, the recurrent strategy is the most reliable method under the condition that short-term prediction models have the ideal 100% accuracy.

But, in most real applications, albeit the short-term models between the two adjacent moments, its prediction performance cannot achieve the perfect 100% due to several uncontrolled factors. Note that, in the direct way, we just need to forecast the APCs, whereas it also requires to predict the MFs (for example, wind speed, which is jointed determined by lots of complicated factors) in the recurrent strategy to be served as the inputs for the subsequent predictions. In such case, the imperfect short-term prediction models must bring about the accumulation and diffusion of errors. Similarly, the recurrent-based compression technologies, due to quantization etc, are always imperfect and lossy, and thus during the compression, what we really save is not $\epsilon_{i\to j}$, but its quantized version of $\tilde{\epsilon}_{i\to j}$. This apparently leads to the accumulation and diffusion of errors. To this end, we acquire the quantified difference $\tilde{\epsilon}_{0\to 1}$ and the associated reconstructed $\tilde{\theta}_1$ based on the known θ_0. Next, we preserve $\tilde{\epsilon}_{0\to 1}$ and compute the quantified error difference $\tilde{\epsilon}_{1\to 2}$ between θ_2 and $P_2(\tilde{\theta}_1)$ since only $\tilde{\theta}_1$ can be obtained during the decompression. Repeat the above steps until the signal has been wholly compressed. On account of the high correlation of two signal values at the neighboring moments, the predicted $P_1(\theta_0)$ is quite close to θ_1 and thus $\epsilon_{0\to 1}$ approaches zero. Therefore, after quantization etc, $\tilde{\epsilon}_{0\to 1}$ also approaches zero and the reconstructed $\tilde{\theta}_1 \approx \theta_1$. Then, we derive that the predicted $P_2(\tilde{\theta}_1)$ is close to θ_2, and $\tilde{\epsilon}_{1\to 2}$ approaches zero and $\tilde{\theta}_2 \approx \theta_2$. Likewise, we can draw the following two results: 1) $\tilde{\theta}_i \approx \theta_i$, where $i > 0$; 2) $\tilde{\epsilon}_{i\to i+1}$ is close to zero. Via the analyses above, we can reasonably assume that the recurrent air quality prediction is a better selection than the direct manner. Apart from the high performance, we are able to derive the predicted air quality indices at the middle moments as well.

In what follows, more concrete and quantified comparison between the direct and recurrent air quality predictions will be presented. We denote the prediction model based on the direct way as $\Upsilon_{j-c} : \mathbf{f}_c \to \mathbf{f}_j$, which is differentiable and predicts $\mathbf{f}_j$ using $\mathbf{f}_c$, where the vector $\mathbf{f}_j = \{f_{j,1}, f_{j,2}, \ldots, f_{j,g}\}$ including g meteorology- and pollution-related parameters. Considering the fact that the prediction accuracy cannot be perfect, the exact expression is $\mathbf{f}_j = \Upsilon_{j-c}(\mathbf{f}_c) + \boldsymbol{\varepsilon}_{j-c}$, where $\boldsymbol{\varepsilon}_{j-c}$ means the error vector. On this basis, we can derive the subsequent equations:

$$\begin{cases} \mathbf{f}_1 = & \Upsilon_1(\mathbf{f}_0) + \boldsymbol{\varepsilon}_1 \\ \mathbf{f}_2 = & \Upsilon_1(\mathbf{f}_1) + \boldsymbol{\varepsilon}_1 \\ & \vdots \\ \mathbf{f}_j = & \Upsilon_1(\mathbf{f}_{j-1}) + \boldsymbol{\varepsilon}_1 \end{cases} . \tag{3}$$

Supposing a vector $\mathbf{z}$, we attain the Taylor series of $\Upsilon_1(\mathbf{z})$ at $\mathbf{z}_0 (= \mathbf{z} + \boldsymbol{\varepsilon})$:

$$\Upsilon_1(\mathbf{z}) = \frac{\Upsilon_1(\mathbf{z}_0)}{0!} + \frac{\Upsilon_1'(\mathbf{z}_0)}{1!}\boldsymbol{\varepsilon} + \frac{\Upsilon_1''(\mathbf{z}_0)}{2!}\boldsymbol{\varepsilon}^2 + \cdots + R(\mathbf{z}) \tag{4}$$

where $\Upsilon_1'(\mathbf{z}_0)$ and $\Upsilon_1''(\mathbf{z}_0)$ are the first- and second-order derivatives; $R(\mathbf{z})$ is an extremely small error term. The value of $\boldsymbol{\varepsilon}$ is small for the short-term prediction model Υ_1, that is to say, $\boldsymbol{\varepsilon}^j$ ($j \geq 2$) is quite close to zero, so we only preserve the former two terms on the right side of (4):

$$\Upsilon_1(\mathbf{z}) \approx \Upsilon_1(\mathbf{z} + \boldsymbol{\varepsilon}) + \Upsilon_1'(\mathbf{z} + \boldsymbol{\varepsilon})\boldsymbol{\varepsilon}. \tag{5}$$

Then, we combine (3) and (5) to derive

$$\begin{aligned} \mathbf{f}_j &= \Upsilon_1(\mathbf{f}_{j-1}) + \boldsymbol{\varepsilon}_1 \\ &= \Upsilon_1(\Upsilon_1(\mathbf{f}_{j-2}) + \boldsymbol{\varepsilon}_1) + \boldsymbol{\varepsilon}_1 \\ &\approx \Upsilon_1^2(\mathbf{f}_{j-2}) - M_{j-2}\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_1 \\ &= \Upsilon_1^2(\Upsilon_1(\mathbf{f}_{j-3}) + \boldsymbol{\varepsilon}_1) - M_{j-2}\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_1 \\ &\approx \Upsilon_1^3(\mathbf{f}_{j-3}) - M_{j-3}\boldsymbol{\varepsilon}_1 - M_{j-2}\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_1 \\ &\quad \vdots \\ &\approx \Upsilon_1^j(\mathbf{f}_0) + \left(1 - \sum_{k=0}^{j-2} M_k\right)\boldsymbol{\varepsilon}_1 \end{aligned} \tag{6}$$

where $M_k = \Upsilon_1^{(j-k-1)'}(\Upsilon_1(\mathbf{f}_k) + \boldsymbol{\varepsilon}_1)$. When forecasting $\mathbf{f}_j$ from $\mathbf{f}_0$ (i.e., $\mathbf{f}_j = \Upsilon_j(\mathbf{f}_0) + \boldsymbol{\varepsilon}_j$), the predicted value is $\Upsilon_j(\mathbf{f}_0)$ and the associated error is $\boldsymbol{\varepsilon}_j$ in the direct way, while in the recurrent-based RAQP model, we obtain the predicted value

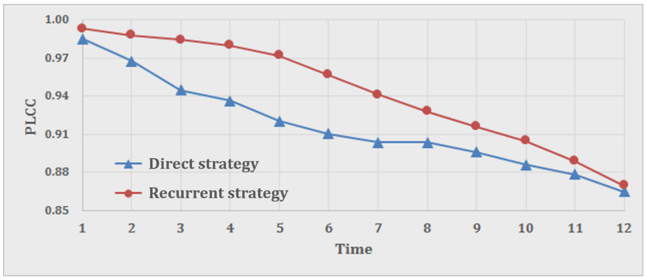


Fig. 3. Comparison of the direct and recurrent strategies for O_3.

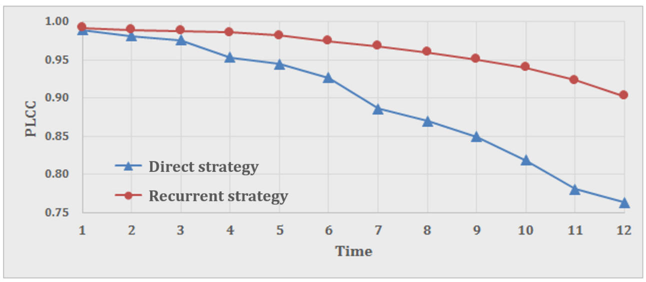


Fig. 4. Comparison of the direct and recurrent ways for PM2.5.

$\Upsilon_1^j(\mathbf{f}_0)$ and the error $(1-\sum_{k=0}^{j-2} M_k)\varepsilon_1$. As thus, comparing the magnitude of ε_j and $(1-\sum_{k=0}^{j-2} M_k)\varepsilon_1$ can lead to the straightforward and reliable result about which one is better between the direct and recurrent strategies. From Fig. 2, we can find that the prediction performance at the T_1 moment is very close to 1, namely ε_1 is extremely small. In this case, we can derive that $|\varepsilon_j|$ is larger than $|(1-\sum_{k=0}^{j-2} M_k)\varepsilon_1|$ and thus the recurrent strategy is superior to the direct manner.

Furthermore, we in this paper introduce noise injection into the feature inputs, which has been widely used for enhancing the generalization capability of the trained regression models [40], [41]. More concretely, we randomly generate 100 noise sets and add them to the training feature set to make the training samples 100 times larger. Note that, on one side, leveraging noised features is able to remarkably increase the number of training samples, and on the other side, it is closer to the real application scenarios because noises must be involved in the measurements of MFs and APCs obtained from instruments. Both the above two sides can raise the regression module's generalization. In addition, it was also found that, due to the use of noised features, the problem of error's accumulation and diffusion can be largely alleviated, so in the recurrent strategy, the error $(1-\sum_{k=0}^{j-2} M_k)\varepsilon_1$ becomes ε_1. According to Fig. 2, $|\varepsilon_1|$ is much less than $|\varepsilon_j|$ and thus the recurrent strategy outperforms the direct way. Akin to Fig. 2, we roughly compare the SVR-based direct prediction model with the recurrent strategy using noised features, i.e., RAQP, in Figs. 3 and 4. Results validate the superiority of RAQP.

III. Results and Discussions

This section will mainly examine the performance of our proposed recurrent-based RAQP model and compare it with the direct strategy and three prevailing air quality prediction models.

A. Experimental Setup

The proposed RAQP model aims to use the current MFs and APCs to infer the hourly estimations of air quality. For training and examining our RAQP model, we collected the hourly records of MFs and APCs at a small village, about 100 km away from Beijing, China. The gathered MFs and APCs (and parts of their units) include time, temperature (°C), relative humidity (%), wind speed (m/s), pressure (hPa), visibility (km), AOT, CO (ppm), NO_2 (ppb), O_3 (ppb), and PM2.5 (μg/m^3). Four APCs to be predicted are CO, NO_2, O_3, and PM2.5. We have successively collected the data over a week, and finally attained 180 h of measurements. In this paper, we apply the two frequently used evaluation measures, root mean square error (RMSE) and PLCC, to check the effectiveness of the proposed RAQP predictor: 1) RMSE measures the prediction consistency, as defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{L}\sum_{l=1}^{L}(a_l - b_l)^2} \tag{7}$$

where a_l and b_l are the predicted and observed values, and L is the number of the elements in one vector; 2) PLCC reflects the prediction accuracy of two vectors, which is defined by

$$\text{PLCC} = \frac{\sum_{l=1}^{L}(a_l-\bar{a})(b_l-\bar{b})}{\sqrt{\sum_{l=1}^{L}(a_l-\bar{a})^2\sum_{l=1}^{L}(b_l-\bar{b})^2}} \tag{8}$$

where $\bar{a}$ and $\bar{b}$ are, respectively, the means of $\mathbf{a}$ and $\mathbf{b}$. A good prediction model is expected to achieve the value of RMSE close to 0, and the value of PLCC close to 1. The competitors consist of three prevailing air quality prediction models. The first Voukantsis model was proposed in [25] by combining two specific computational intelligence methods, separately principal component analysis and artificial neural networks. The second Vlachogianni method was developed in [26] with the stepwise multiple linear regression. The last Kaboodvandpour predictor was devised in [27] based on the adaptive neuro-fuzzy inference system. In the comparison to be illustrated later, all the models use the same features for air quality prediction.

B. Performance Evaluation

For the direct strategy and the three air quality prediction models compared, we randomly separate the entire dataset into two teams. One team contains 80% data for training and the other contains the rest 20% data for testing. We repeat the aforementioned process 100 times and compute the PLCC and RMSE indices to measure the average prediction performance from the T_1 to T_{12} moments, as illustrated in Tables I and II. When inferring the prediction performance of the recurrent-based RAQP model, we first use the noised features, which are created by adding 100 randomly generated noise sets, to learn the 1-h prediction model, and then repeatedly use the 1-h prediction model n times to infer air quality at the T_n moment (i.e., after n hours). The results of the proposed RAQP predictor can be also found in Tables I and II. For easy comparison, we highlight the first- and second-rank models with boldface and underline, respectively.

TABLE I
PLCC Comparison Among the Direct Way, Our RAQP Model, and Three Popular Predictors

APC	Model	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}
CO	Direct manner	0.9763	0.9712	0.9666	0.9670	**0.9639**	**0.9622**	**0.9558**	**0.9452**	**0.9240**	**0.9222**	**0.9187**	**0.9157**
	RAQP (Pro.)	**0.9931**	**0.9860**	**0.9780**	**0.9711**	0.9615	0.9478	0.9303	0.9049	0.8728	0.8325	0.7546	0.6128
	Voukantsis	0.9147	0.8973	0.8683	0.8382	0.8013	0.7578	0.7368	0.6883	0.6151	0.5730	0.5515	0.5498
	Vlachogianni	0.9792	0.9620	0.9471	0.9423	0.9224	0.9105	0.8981	0.8770	0.8527	0.8329	0.8134	0.8064
	Kaboodvandpour	0.9329	0.8868	0.8812	0.8810	0.8768	0.8737	0.8533	0.8384	0.8366	0.7833	0.7506	0.7457
NO_2	Direct manner	0.9884	0.9807	0.9664	0.9533	0.9472	0.9300	0.8993	0.8882	0.8758	0.8477	0.8298	0.8228
	RAQP (Pro.)	**0.9950**	**0.9920**	**0.9886**	**0.9852**	**0.9804**	**0.9728**	**0.9647**	**0.9552**	**0.9446**	**0.9301**	**0.9104**	**0.8849**
	Voukantsis	0.9440	0.9183	0.8812	0.8459	0.8182	0.7789	0.7094	0.6788	0.6480	0.6087	0.5037	0.5008
	Vlachogianni	0.9904	0.9800	0.9676	0.9536	0.9446	0.9262	0.8978	0.8854	0.8654	0.8434	0.8177	0.8158
	Kaboodvandpour	0.9031	0.9029	0.9000	0.8928	0.8793	0.8545	0.8159	0.7936	0.7878	0.7183	0.7137	0.6677
O_3	Direct manner	0.9851	0.9675	0.9448	0.9365	0.9202	0.9102	0.9036	0.9035	0.8959	0.8859	0.8784	0.8647
	RAQP (Pro.)	**0.9933**	**0.9881**	**0.9845**	**0.9800**	**0.9719**	**0.9568**	**0.9412**	**0.9280**	**0.9160**	**0.9048**	**0.8886**	**0.8694**
	Voukantsis	0.8823	0.8566	0.8263	0.8037	0.7590	0.7060	0.6292	0.6022	0.5287	0.4769	0.3889	0.3674
	Vlachogianni	0.9834	0.9617	0.9390	0.9271	0.9173	0.9031	0.8975	0.8934	0.8775	0.8622	0.8403	0.8305
	Kaboodvandpour	0.7701	0.7614	0.7352	0.7288	0.6824	0.6717	0.6630	0.6395	0.6340	0.6145	0.5528	0.5393
PM2.5	Direct manner	0.9893	0.9814	0.9762	0.9531	0.9445	0.9267	0.8861	0.8696	0.8490	0.8183	0.7808	0.7628
	RAQP (Pro.)	**0.9921**	**0.9896**	**0.9880**	**0.9857**	**0.9820**	**0.9748**	**0.9678**	**0.9597**	**0.9504**	**0.9397**	**0.9234**	**0.9027**
	Voukantsis	0.9378	0.9283	0.9136	0.8813	0.8627	0.8418	0.8005	0.7695	0.7467	0.7091	0.6656	0.6260
	Vlachogianni	0.9893	0.9794	0.9719	0.9590	0.9471	0.9321	0.9084	0.8883	0.8687	0.8484	0.8189	0.8128
	Kaboodvandpour	0.8289	0.8069	0.7680	0.7558	0.7404	0.7306	0.6953	0.6717	0.6546	0.6398	0.6143	0.5995

Note: We bold the best-performing one.

TABLE II
RMSE Comparison Across the Direct Way, our RAQP Model, and Three Popular Predictors

APC	Model	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}
CO	Direct manner	0.1882	0.2067	0.2202	0.2269	**0.2332**	**0.2372**	**0.2559**	**0.2836**	**0.3186**	**0.3261**	**0.3395**	**0.3412**
	RAQP (Pro.)	**0.0998**	**0.1425**	**0.1784**	**0.2042**	0.2343	0.2709	0.3108	0.3604	0.4169	0.4835	0.6128	0.8763
	Voukantsis	0.3763	0.4053	0.4385	0.4788	0.5369	0.5776	0.6091	0.6239	0.6674	0.6888	0.7155	0.7051
	Vlachogianni	0.1776	0.2329	0.2717	0.2968	0.3290	0.3530	0.3899	0.4137	0.4434	0.4701	0.4931	0.4991
	Kaboodvandpour	0.6303	0.6561	0.6582	0.6596	0.6663	0.6744	0.6799	0.6987	0.7258	0.7465	0.7656	0.7746
NO_2	Direct manner	4.1528	5.3493	6.8445	8.0506	8.8996	10.252	11.544	12.377	13.600	14.761	15.534	15.917
	RAQP (Pro.)	**2.6389**	**3.3312**	**3.9785**	**4.5397**	**5.2179**	**6.1318**	**6.9807**	**7.8607**	**8.7273**	**9.7764**	**11.021**	**12.438**
	Voukantsis	9.8196	11.279	13.644	14.780	15.975	17.251	18.996	19.866	20.634	21.848	22.709	23.282
	Vlachogianni	3.7322	5.2322	6.7695	7.9972	8.8784	10.197	11.497	12.424	13.616	14.529	15.173	15.621
	Kaboodvandpour	16.630	17.002	17.889	18.435	18.529	18.655	18.788	19.190	19.229	19.476	20.274	21.485
O_3	Direct manner	2.5714	3.8535	4.9092	5.3734	6.0635	6.2456	6.3914	6.4499	6.5306	6.7123	7.2334	7.8643
	RAQP (Pro.)	**1.7030**	**2.2789**	**2.6188**	**2.9885**	**3.5571**	**4.4089**	**5.1343**	**5.6756**	**6.1356**	**6.5564**	**7.1286**	**7.7436**
	Voukantsis	7.2805	7.8656	8.7747	9.2737	10.093	10.794	11.551	12.150	12.890	13.266	13.778	13.977
	Vlachogianni	2.6477	4.1064	5.2081	5.6544	6.0855	6.5257	6.5790	6.7806	7.1891	7.5889	8.2691	8.5902
	Kaboodvandpour	10.718	11.098	11.117	11.578	11.846	11.917	11.953	11.955	12.134	12.143	12.290	12.677
PM2.5	Direct manner	8.6753	11.057	13.222	17.037	20.232	23.374	26.830	30.015	31.824	34.477	35.957	38.534
	RAQP (Pro.)	**7.1740**	**8.2678**	**8.9187**	**9.7672**	**10.955**	**12.943**	**14.580**	**16.249**	**17.950**	**19.711**	**22.111**	**24.785**
	Voukantsis	22.470	23.662	25.492	29.234	30.698	33.309	36.485	38.155	40.516	41.993	44.050	46.663
	Vlachogianni	8.9351	11.758	13.683	16.438	18.826	21.771	24.180	26.648	28.933	30.976	32.929	34.584
	Kaboodvandpour	38.353	39.833	42.351	42.364	43.040	43.244	45.505	46.091	46.861	47.092	47.666	49.436

Note: We bold the best-performing one.

From Tables I and II, we are able to draw the following two conclusions. First, we make a comparison between the direct strategy and the proposed recurrent-based RAQP predictor. In most situations, our RAQP model has led to the greater performance than the direct way. To specify, when inferring the concentration of CO, the RAQP predictor performs better from T_1 to T_4, while as for NO_2, O_3, and PM2.5, the RAQP model is constantly superior to the direct strategy from the T_1 to T_{12} moments. Second, the proposed RAQP predictor is compared with the three state-of-the-art air quality prediction models. Our RAQP predictor has also achieved encouraging results. Particularly, the proposed RAQP model is better than the three competitors for predicting the concentrations of four air pollutants constantly, except that it is only inferior to the Vlachogianni model for CO from T_{10} to T_{12}. The reason that the recurrent-based RAQP predictor is not always the best is very possible due to the fact that the 1-h prediction model is not perfect and therefore the accumulation and diffusion of errors still appears. In other words, we may further improve the performance of our RAQP model through introducing the better-performance 1-h prediction model.

Furthermore, we more concern the prediction performance during the next 3–4 h, since as for daily trips, this period is sufficient to make us go back homes or offices once we know

TABLE III
STATISTICAL SIGNIFICANCE COMPARISON BETWEEN THE DIRECT AND RECURRENT STRATEGIES

APC	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}
CO	**+1**	**+1**	**+1**	**+1**	0	−1	−1	−1	−1	−1	−1	−1
NO_2	**+1**	**+1**	**+1**	**+1**	**+1**	**+1**	**+1**	**+1**	**+1**	**+1**	**+1**	**+1**
O_3	**+1**	**+1**	**+1**	**+1**	**+1**	**+1**	**+1**	**+1**	**+1**	**+1**	**+1**	0
PM2.5	**+1**	**+1**	**+1**	**+1**	**+1**	**+1**	**+1**	**+1**	**+1**	**+1**	**+1**	**+1**

TABLE IV
PLCC COMPARISON BETWEEN THE ORIGINAL AND RECURRENT-BASED VLACHOGIANNI MODELS

APC	Model	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}	T_{11}	T_{12}
CO	Vlachogianni	0.9792	0.9620	0.9471	0.9423	0.9224	0.9105	0.8981	0.8770	0.8527	**0.8329**	**0.8134**	**0.8064**
	R-Vlachogianni	**0.9802**	**0.9672**	**0.9542**	**0.9432**	**0.9294**	**0.9151**	**0.8997**	**0.8781**	**0.8549**	0.8321	0.8084	0.7860
NO_2	Vlachogianni	0.9904	0.9800	0.9676	0.9536	**0.9446**	**0.9262**	0.8978	**0.8854**	**0.8654**	**0.8434**	**0.8177**	**0.8158**
	R-Vlachogianni	**0.9915**	**0.9814**	**0.9695**	**0.9555**	0.9393	0.9194	**0.8979**	0.8750	0.8502	0.8243	0.7980	0.7720
O_3	Vlachogianni	0.9834	0.9617	0.9390	**0.9271**	**0.9173**	**0.9031**	**0.8975**	**0.8934**	**0.8775**	**0.8622**	**0.8403**	**0.8305**
	R-Vlachogianni	**0.9857**	**0.9644**	**0.9420**	0.9200	0.8973	0.8741	0.8561	0.8410	0.8265	0.8182	0.8092	0.8018
PM2.5	Vlachogianni	0.9893	0.9794	0.9719	0.9590	0.9471	0.9321	0.9084	0.8883	0.8687	0.8484	0.8189	**0.8128**
	R-Vlachogianni	**0.9895**	**0.9814**	**0.9729**	**0.9627**	**0.9490**	**0.9329**	**0.9150**	**0.8965**	**0.8748**	**0.8539**	**0.8320**	0.8079

Note: We highlight the best-performing one.

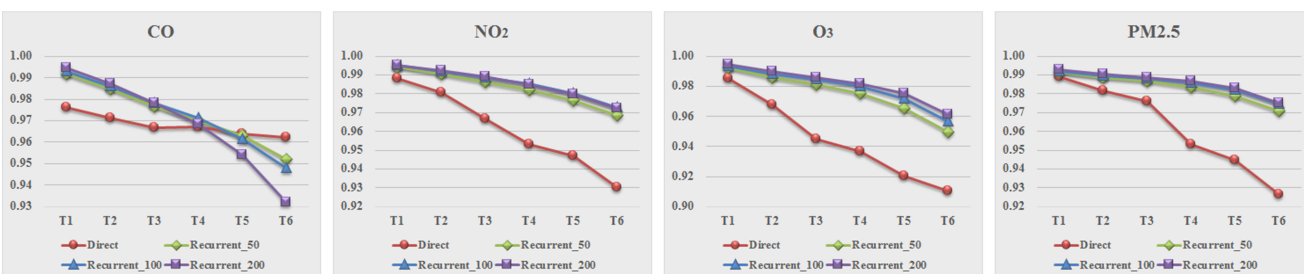


Fig. 5. Variations of influences on different air pollutants with the number of randomly generated noise sets changed.

the air quality will go bad. Concretely, we just pay attention to the PLCC index for O_3 and PM2.5, which are the two most significant air pollutants concerned by the governments during recent years, and interested readers can compute the variations of RMSE and other air pollutants. For O_3, the proposed RAQP prediction model has independently resulted in about 4.2% and 4.6% relative performance gains than the second-ranking model at the T_3 and T_4 moments. As for PM2.5, the relative gains between our RAQP model and the second-performer is around 1.2% and 2.8% at T_3 and T_4, respectively.

C. Statistical Comparison

We also implement the statistical significance comparison. *T*-test is used to examine whether two sets of data are significantly different from each other [42]. The *t*-test is perhaps the most commonly used test under the condition that the test statistic follows a normal distribution. We apply the *t*-test to the PLCC indices, which are acquired from the 100 times 80% train to 20% test trials, of the direct and recurrent strategies for hourly predictions of four APCs during the next 12 h. Tables III lists the results of statistical significance comparison. The null hypothesis means that the average PLCC value for one model is equal to that for another model with a confidence of 95%. The value of "0" means that the two strategies are statistically equivalent to each other, the value of "+1" means that the recurrent strategy is statistically superior to the direct strategy, and the value of "−1" means that the recurrent strategy is statistically inferior to the direct strategy. We bold all the "+1" for the readers' conveniences. One can see that our RAQP model has delivered statistically higher performance in the absolute majority of conditions.

D. Generality

The proposed RAQP predictor also provides a general-purpose framework that is applicable to improving the high-accuracy air quality prediction models toward better performance. According to the analyses mentioned above, the basic premise of using the recurrent framework lies in the high performance of the 1-h prediction model. Comparing the results reported in Tables I and II, the Vlachogianni model is accordingly selected to check the generality of the proposed recurrent framework. To specify, the 1-h Vlachogianni model

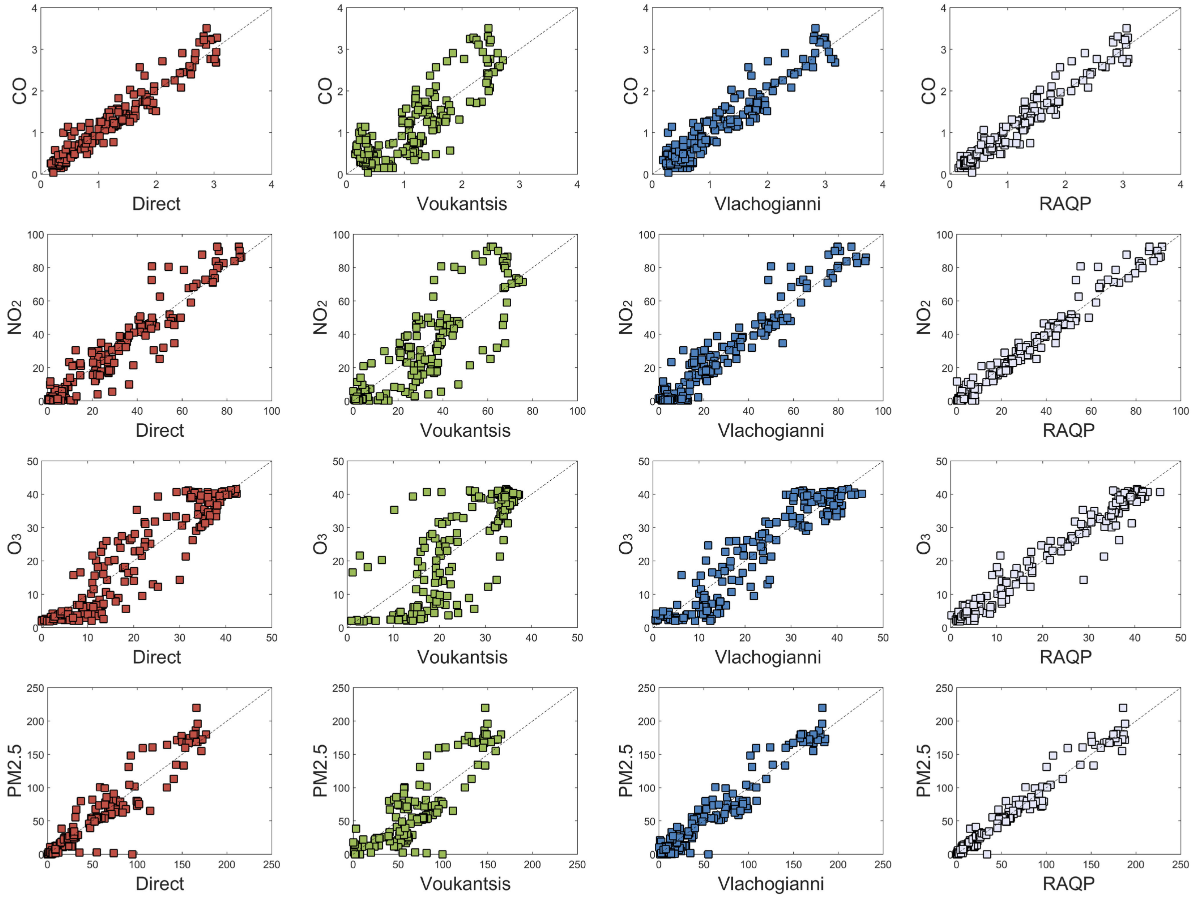


Fig. 6. Scatter plots of concentrations of different air pollutants versus predictions of four competing models at the T_4 moment.

is established based on the noised features, followed by being repeatedly used n times to forecast the air quality at the T_n moment (i.e., after n hours). We tabulate the PLCC results of the recurrent-based Vlachogianni (R-Vlachogianni) model in Table IV, and simultaneously we also list the result of the original Vlachogianni model and bold the better one for easy comparison. It can be found that the recurrent framework is capable of enhancing the Vlachogianni model's accuracy, especially for the short- and mid-term predictions such as at the T_2 and T_3 moments. Additionally, we find that, for the long-term prediction, the R-Vlachogianni model does not work well, and this is mainly due to the reason that the insufficiently high performance of the 1-h Vlachogianni model speeds up the accumulation and diffusion of errors.

E. Influences of Parameters

How to learn reliably has long been a critical open problem. In [43], the authors have put forward a differential privacy-based thresholdout technology to resolve the overfitting problem caused by the nonreusable holdout. Inspired by several parts in the above publication, the proposed RAQP model is proposed by introducing 100 randomly generated noise sets to improve the generalizability when training the model. The influence of the number of noise sets on the correlation performance is significant, since it will reduce the computational complexity and raise the implementation speed if less randomly produced noise sets are exploited. With this concern, this paper checks the variations of impacts when the number is assigned as 50, 100, and 200. We plot the four scatter plots to illustrate the results, as displayed in Fig. 5. One can readily find that, as the number grows from 50 to 200, the performance is increased but the change is not evident. So we may use a small amount of randomly created noise sets in real applications. Furthermore, it is worthy to notice that our work dominantly concentrates on providing a heuristic solution to air quality prediction, not focuses on all the implementation details such as the number of the added noise sets. The future work might be devoted to exploring how to determine it faithfully.

F. Visualized Comparison

The scatter plot is a very important and direct manner for performance comparison. Through scatter plots, the readers can easily observe which model is superior to others and why other models performed not so well. Hence, we illustrate the scatter plots of concentrations of four air pollutants (i.e., CO, NO_2, O_3 and PM2.5) versus predictions of four competing models at the T_4 moment, as shown in Fig. 6. Via visualized straightforward comparison, we can obviously find that the sample points of the proposed RAQP predictor present higher convergence and linearity than other prediction models tested. This reveals that our predictor is able to yield more consistent predictions in line with the truth values.

G. Discussions

We in this paper introduce a novel recurrent strategy, which has been extensively used in many industrial applications such as compression of acoustic, video, and power signals, into the prediction of concentrations of air pollutants. We use 11 relevant features, encompassing time, temperature, relative humidity, wind speed, pressure, visibility, AOT, CO, NO_2, O_3, and PM2.5, to separately predict each of above 11 variables at the next moment. Based on the predicted results of APCs at some time later, the standard equation can be used to combine them and derive the overall quality score [44]. By recurrently implementing the aforesaid procedure, we are able to predict the APCs and air quality index several hours later. Experimental results prove the effectiveness of our recurrent-based RAQP model as compared with the direct strategy and state-of-the-art air quality prediction models. By comparison with the early studies, the two major contributions have been made in our work. First, to our best knowledge, we are the first to apply the recurrent strategy to air quality prediction, which provides not only a high-accuracy RAQP predictor, but also a general-purpose framework applicable to improving the performance of existing air quality predictors. Second, this paper first introduces the noised features into the air quality estimation, which simultaneously enhances the generalization of models and addresses the problem of error accumulation. However, the proposed RAQP model is not always the best since there still exists a gap between the 1-h prediction model and the ideal 100% performance. In the future, deep-based unsupervised and supervised machine learning tools may be included to better reveal the nonlinear relationship between the input MFs and APCs and output APCs, particularly modifying the 1-h prediction model Υ_1 and lowering ε_1 and thus improving our proposed RAQP predictor's accuracy.

IV. Conclusion

In this paper, we have investigated into an emerging and urgent problem−air quality prediction. Considering the nonlinear relationship between the MFs and APCs at the present moment and the air quality indices several hours later, a heuristic recurrent-based RAQP model has been proposed by recurrently using the 1-h prediction model toward mid- and long-term predictions. The proposed recurrent strategy fully exploits the advantage of high correlations between the meteorology- and pollution-related parameters during short time intervals to build a 1-h prediction model, and meanwhile, avoids the disadvantage of their weak correlations as the time interval increases to large through using the recurrent strategy. Results of experiments demonstrate the effectiveness of our proposed RAQP model as compared with the direct strategy and state-of-the-art air quality predictors, and furthermore, confirm the generality ability of the proposed recurrent framework. In addition to the high performance, the predicted air quality indices at the middle moments can be also derived in the recurrent strategy. The future work will be devoted to improving the RAQP predictor by modifying the 1-h prediction model based on ensemble learning and deep learning.

References

[1] Y. Song *et al.*, "Source apportionment of PM2.5 in Beijing by positive matrix factorization," *Atmos. Environ.*, vol. 40, no. 8, pp. 1526–1537, Mar. 2006.

[2] B. Lv, Y. Hu, H. H. Chang, A. G. Russell, and Y. Bai, "Improving the accuracy of daily PM2.5 distributions derived from the fusion of ground-level measurements with aerosol optical depth observations, a case study in north China," *Environ. Sci. Technol.*, vol. 50, no. 9, pp. 4752–4759, Apr. 2016.

[3] J. Wang and S. A. Christopher, "Intercomparison between satellite-derived aerosol optical thickness and PM2.5 mass: Implications for air quality studies," *Geophys. Res. Lett.*, vol. 30, no. 21, Nov. 2003.

[4] Y. Liu, J. A. Sarnat, V. Kilaru, D. J. Jacob, and P. Koutrakis, "Estimating ground-level PM2.5 in the eastern United States using satellite remote sensing," *Environ. Sci. Technol.*, vol. 39, no. 9, pp. 3269–3278, Mar. 2005.

[5] A. V. Donkelaar, R. V. Martin, and R. J. Park, "Estimating ground-level PM2.5 using aerosol optical depth determined from satellite remote sensing," *J. Geophys. Res.: Atmospheres*, vol. 111, no. D21, Nov. 2006.

[6] J. A. Engel-Cox, C. H. Holloman, B. W. Coutant, and R. M. Hoff, "Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality," *Atmos. Environ.*, vol. 38, no. 16, pp. 2495–2509, May 2004.

[7] D. E. Abbey, B. E. Ostro, G. Fraser, T. Vancuren, and R. J. Burchette, "Estimating fine particulates less than 2.5 microns in aerodynamic diameter (PM2.5) from airport visibility data in California," *J. Exposure Anal Environ. Epidemiol.*, vol. 5, no. 2, pp. 161–180, Apr. 1995.

[8] M. Zheng *et al.*, "Seasonal trends in PM2.5 source contributions in Beijing, China," *Atmos. Environ.*, vol. 39, no. 22, pp. 3967–3976, Jul. 2005.

[9] J.-L. Wang *et al.*, "Quantitative relationship between visibility and mass concentration of PM2.5 in Beijing," *J. Environ. Sci.*, vol. 18, no. 3, pp. 475–481, Jun. 2006.

[10] R. M. Harrison, A. R. Deacon, and M. R. Jones, "Sources and processes affecting concentrations of PM10 and PM2.5 particulate matter in Birmingham (U.K.)," *Atmos. Environ.*, vol. 31, no. 24, pp. 4103–4117, Dec. 1997.

[11] G. M. Marcazzan, S. Vaccaro, G. Valli, and R. Vecchi, "Characterisation of PM10 and PM2.5 particulate matter in the ambient air of Milan (Italy)," *Atmos. Environ.*, vol. 35, no. 27, pp. 4639–4650, Sep. 2001.

[12] C.-H. Huang and C.-Y. Tai, "Relative humidity effect on PM2.5 readings recorded by collocated beta attenuation monitors," *Environ. Eng. Sci.*, vol. 25, no. 7, pp. 1079–1090, Sep. 2008.

[13] C.-X. Zhao, Y.-Q. Wang, Y.-J. Wang, H.-L. Zhang, and B.-Q. Zhao, "Temporal and spatial distribution of PM2.5 and PM10 pollution status and the correlation of particulate matters and meteorological factors during winter and spring in Beijing," *Environ. Sci.*, vol. 35, no. 2, pp. 418–427, Feb. 2014.

[14] J. Wang and S. Ogawa, "Effects of meteorological conditions on PM2.5 concentrations in Nagasaki, Japan," *Int. J. Environ. Res. Public Health*, vol. 12, no. 8, pp. 9089–9101, Aug. 2015.

[15] Y. Wang, Q. Ying, J. Hu, and H. Zhang, "Spatial and temporal variations of six criteria air pollutants in 31 provincial capital cities in China during 2013-2014," *Environ. Int.*, vol. 73, pp. 413–422, Dec. 2014.

[16] G. B. Raga and L. Le Moyne, "On the nature of air pollution dynamics in Mexico City−I. Nonlinear analysis," *Atmos. Environ.*, vol. 30, no. 23, pp. 3987–3993, Dec. 1996.

[17] V. Valverde, M. T. Pay, and J. M. Baldasano, "Circulation-type classification derived on a climatic basis to study air quality dynamics over the Iberian Peninsula," *Int. J. Climatol.*, vol. 35, no. 10, pp. 2877–2897, Aug. 2015.

[18] C. Chemel *et al.*, "Application of chemical transport model CMAQ to policy decisions regarding PM2.5 in the UK," *Atmos. Environ.*, vol. 82, pp. 410–417, Jan. 2014.

[19] H. J. S. Fernando *et al.*, "Forecasting PM10 in metropolitan areas: efficacy of neural networks," *Environ. Pollut.*, vol. 163, pp. 62–67, Apr. 2012.

[20] M. Elbayoumi, N. A. Ramli, N. F. F. M. Yusof, A. S. B. Yahaya, W. A. Madhoun, and A.Z. Ul-Saufie, "Multivariate methods for indoor PM10 and PM2.5 modelling in naturally ventilated schools buildings," *Atmos. Environ.*, vol. 94, pp. 11–21, Sep. 2014.

[21] H. Niska, T. Hiltunen, A. Karppinen, J. Ruuskanen, and M. Kolehmainen, "Evolving the neural network model for forecasting air pollution time series," *Eng. Appl. Artif. Intell.* vol. 17, no. 2, pp. 159–167, Mar. 2004.

[22] G. Corani, "Air quality prediction in Milan: Feed-forward neural networks, pruned neural networks and lazy learning," *Ecol. Model.* vol. 185, no. 2, pp. 513–529, Jul. 2005.

[23] S. M. S. Nagendra and M. Khare, "Artificial neural network approach for modelling nitrogen dioxide dispersion from vehicular exhaust emissions," *Ecol. Model.* vol. 190, no. 1, pp. 99–115, Jan. 2006.

[24] J. B. Ordieres, E. P. Vergara, R. S. Capuz, and R. E. Salazar, "Neural network prediction model for fine particulate matter (PM2.5) on the US−Mexico border in El Paso (Texas) and Ciudad Juárez (Chihuahua)," *Environ. Model. Softw.* vol. 20, no. 5, pp. 547–559, May 2005.

[25] D. Voukantsis, K. Karatzas, J. Kukkonen, TeemuRäsänen, A. Karppinen, and M. Kolehmainen, "Intercomparison of air quality data using principal component analysis, and forecasting of PM_{10} and $PM_{2.5}$ concentrations using artificial neural networks, in Thessaloniki and Helsinki," *Sci. Total Environ.*, vol. 409, no. 7, pp. 1266–1276, Mar. 2011.

[26] A. Vlachogianni, P. Kassomenos, A. Karppinen, S. Karakitsios, and J. Kukkonen, "Evaluation of a multiple regression model for the forecasting of the concentrations of NO_x and PM_{10} in Athens and Helsinki," *Sci. Total Environ.*, vol. 409, no. 8, pp. 1559–1571, Mar. 2011.

[27] S. Kaboodvandpour, J. Amanollahi, S. Qhavami, and B. Mohammadi, "Assessing the accuracy of multiple regressions, ANFIS, and ANN models in predicting dust storm occurrences in Sanandaj, Iran," *Nat. Hazards*, vol. 78, no. 2, pp. 879–893, Apr. 2015.

[28] C-C. Chang and C-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intel. Syst. Technol.*, vol. 2, no. 3, Apr. 2011, Art. no. 27.

[29] G. Wang and J. Jiao, "A kernel least squares based approach for nonlinear quality-related fault detection," *IEEE Trans. Ind. Electron.*, vol. 64, no. 4, pp. 3195–3204, Apr. 2017.

[30] P. Tang and T. W. S. Chow, "Wireless sensor-networks conditions monitoring and fault diagnosis using neighborhood hidden conditional random field," *IEEE Trans. Ind. Informat.*, vol. 12, no. 3, pp. 933–940, Jun. 2016.

[31] Y.-L. Wang and Q.-L. Han, "Network-based fault detection filter and controller coordinated design for unmanned surface vehicles in network environments," *IEEE Trans. Ind. Informat.*, vol. 12, no. 5, pp. 1753–1765, Oct. 2016.

[32] T. Zhao, J. Wang, Z. Wang, and C. W. Chen, "SSIM-based coarse-grain scalable video coding," *IEEE Trans. Broadcasting*, vol. 61, no. 2, pp. 210–221, Jun. 2015.

[33] J. Wang, A. Rehman, K. Zeng, S. Wang, and Z. Wang, "Quality prediction of asymmetrically distorted stereoscopic 3D images," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3400–3414, Nov. 2015.

[34] J. Wang, S. Wang, and Z. Wang, "Asymmetrically compressed stereoscopic 3D videos: Quality assessment and rate-distortion performance evaluation," *IEEE Trans. Image Process.*, vol. 26 no. 3, pp. 1330–1343, Mar. 2017.

[35] K. Gu, L. Li, H. Lu, X. Min, and W. Lin, "A fast reliable image quality predictor by fusing micro- and macro-structures," *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 3903–3912, May 2017.

[36] W. Xiong, X. Hu, and T. Jiang, "Measurement and characterization of link quality for IEEE 802.15. 4-compliant wireless sensor networks in vehicular communications," *IEEE Trans. Ind. Informat.*, vol. 12, no. 5, pp. 1702–1713, Oct. 2016.

[37] L. Zhao, G. Diao, and Y. Yao, "A dynamic process adjustment method based on residual prediction for quality improvement," *IEEE Trans. Ind. Informat.*, vol. 12, no. 1, pp. 41–50, Feb. 2016.

[38] Z. Xiao, H. B. Lim, and L. Ponnambalam, "Participatory sensing for smart cities: A case study on transport trip quality measurement," *IEEE Trans. Ind. Informat.*, vol. 13, no. 2, pp. 759–770, Apr. 2017.

[39] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.

[40] K. Matsuoka, "Noise injection into inputs in back-propagation learning," *IEEE Trans. Syst., Man, Cybern.*, vol. 22, no. 3, pp. 436–440, May 1992.

[41] N. Morales, L. Gu, and Y. Gao, "Adding noise to improve noise robustness in speech recognition," *Interspeech*, pp. 930–933, 2007.

[42] D. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures.* London, U.K.: Chapman & Hall, 2004.

[43] C. Dwork, V. Feldman, M. Hardt, T. Pitassi, O. Reingold, and A. Roth, "The reusable holdout: Preserving validity in adaptive data analysis," *Science*, vol. 349, no. 6248, pp. 636–638, Aug. 2015.

[44] 2018. [Online]. https://en.wikipedia.org/wiki/Air_quality_index

Ke Gu received the B.S. and Ph.D. degrees in electronic engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009 and 2015, respectively.

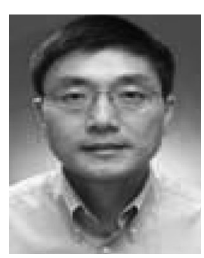
He is an Associate Professor at the Beijing University of Technology, Chaoyang, China. He is currently the Associated Editor for IEEE Access, and is the Reviewer for 20 top SCI journals. His research interests include image processing, environmental perception, quality assessment, machine learning, and big data analysis.

Dr. Gu received the Best Paper Award at the IEEE International Conference on Multimedia and Expo in 2016, and received the excellent Ph.D. thesis award from the Chinese Institute of Electronics in 2016. He was the leading special session organizer in VCIP2016 and ICIP2017, and serves as the Guest Editor for the *Digital Signal Processing* journal.

Jun-Fei Qiao (M'11) received the B.E. and M.E. degrees in control engineer from Liaoning Technical University, Fuxin, China, in 1992 and 1995, respectively, and the Ph.D. degree from Northeast University, Shenyang, China, in 1998.

He was a Postdoctoral Fellow with the School of Automatics, Tianjin University, Tianjin, China, from 1998 to 2000. He joined the Beijing University of Technology, Beijing, China, where he is currently a Professor. He is also the Director of the Intelligence Systems Laboratory. His current research interests include neural networks, intelligent systems, self-adaptive/learning systems, and process control systems.

Prof. Qiao is a member of the IEEE Computational Intelligence Society. He is a Reviewer for more than 20 international journals, such as the IEEE TRANSACTIONS ON FUZZY SYSTEMS and the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS.



Weisi Lin (F'16) received the Ph.D. degree from Kings College London, London, U.K.

He is currently a Full Professor with the School of Computer Engineering, Nanyang Technological University, Singapore. His research interests include image processing, visual quality evaluation, and perception-inspired signal modeling, with more than 340 refereed papers published in international journals and conferences.

Dr. Lin is a Fellow of Institution of Engineering Technology and an Honorary Fellow of the Singapore Institute of Engineering Technologists. He has been on the Editorial Board of the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON MULTIMEDIA (2011–2013), the IEEE SIGNAL PROCESSING LETTERS, and the *Journal of Visual Communication and Image Representation*. He has been elected as an APSIPA Distinguished Lecturer (2012/13). He served as a Technical-Program Chair for Pacific-Rim Conference on Multimedia 2012, the IEEE International Conference on Multimedia and Expo 2013, and the International Workshop on Quality of Multimedia Experience 2014.