A Nonparametric Online Model for Air Quality Prediction

Vitor Guizilini[1] and **Fabio Ramos**[1,2]

[1]National ICT Australia (NICTA),

[2]School of Information Technologies, The University of Sydney, Australia.

Abstract

We introduce a novel method for the continuous online prediction of particulate matter in the air (more specifically, PM10 and PM2.5) given sparse sensor information. A nonparametric model is developed using Gaussian Processes, which eschews the need for an explicit formulation of internal – and usually very complex – dependencies between meteorological variables. Instead, it uses historical data to extrapolate pollutant values both spatially (in areas with no sensor information) and temporally (the near future). Each prediction also contains a respective variance, indicating its uncertainty level and thus allowing a probabilistic treatment of results. A novel training methodology (Structural Cross-Validation) is presented, which preserves the spatio-temporal structure of available data during the hyperparameter optimization process. Tests were conducted using a real-time feed from a sensor network in an area of roughly 50×80 km, alongside comparisons with other techniques for air pollution prediction. The promising results motivated the development of a smartphone applicative and a website, currently in use to increase the efficiency of air quality monitoring and control in the area.

Introduction

Real-Time Air Quality Forecast (RT-AQF), although a fairly new discipline in atmospheric sciences, represents one of the most far-reaching developments and practical applications of science and engineering (Zhang et al. 2012a). Numerous studies (Phalen and Phalen 2011; Greenbaum et al. 2001) show that even acute (short-term) exposure to high level of air pollutants may pose serious health consequences, such as eye irritation, difficulty breathing, pulmonary, cardio-vascular health effects and premature death. Chronic (long-term) exposure may lead to cancer, premature death and damage to the body's immune, neurological, reproductive and respiratory systems. The US. Environmental Protection Agency (EPA) has set air quality standards for six air pollutants: sulfur dioxide (SO_2), nitrogen dioxide (NO_2), carbon monoxide (CO), ozone (O_3), lead (Pb) and particulate matter with aerodynamic diameters less than or equal to 10 μm

(PM10) and 2.5 μm (PM2.5). Despite these efforts, air pollution is estimated to kill 3 million people worldwide, and PM pollution in particular has been directly linked to excess deaths in many countries.

The first RT-AQF techniques were developed in the 1970s, based largely on empirical or statistically trained models (McCollister and Wilson 1975; Wolff and Lioy 1978). The sophistication level of such models increased considerably over time to address challenges, such as non-linearity (Ryan 1995). The development of 3-D numerical air quality models (AQMs) on urban, regional and global scales (Zhang et al. 2012a) also led to significant performance improvements, introducing systems that account for meteorology, emissions, chemistry and removal processes (Zhang 2008; Russel and Dennis 2000). Another performance leap was due to the first real-time data repositories (U.S. EPA 2009), providing effective platforms for continuously communicating air quality conditions and forecasts to the public via the internet and other medias.

Broadly speaking, techniques for RT-AQF can be divided into three categories (Wayland 2002):

Criteria Schemes: Employ rules-of-thumb or thresholds to predict the future state of air quality. For example, if the temperature forecast if above 30^oC, then a photochemical smog event will occur. Another example, often surprisingly robust, is persistence, which means assuming that the air quality tomorrow will be similar to today's (U.S. EPA 2003).

Deterministic methods: Attempt to simulate the underlying chemical and physical process of air pollution production. Usually requires a lot of computational power, and takes into consideration variables such as meteorological conditions, gaseous and particle emissions, chemical reactions and nature of surface (Zhang et al. 2012a; Zhang 2008),.

Nonparametric models: An increasingly popular approach that makes predictions based on patterns in past meteorological, traffic and air quality data, using regression or neural network systems. For example: multiple non-linear regression relations for radiation and temperature might be use to predict peak ozone concentrations (Athanasiadis et al. 2003; Marchant and Ramos 2012; Athanasiadis and Mitkas 2004).

A detailed list of different methods currently used throughout the world can be found in (Cope and Hess

2005). Current state-of-the-art (Zhang et al. 2012b) include 3-D online-coupled meteorology-chemistry models, based on global and regional CTMs (Chemical Transport Models) that enable a better understanding of mechanisms for pollutant formation and development of emission control strategies. Extension to these models include their coupling with urban models (traffic and/or local pollutant dispersion) or a CFD model for urban/local scale applications.

The deterministic approaches rely on parametric models that must be carefully constructed by experts according to which phenomenon is being studied. A balance must be reached between complexity and computational power, which imposes a limit to how many variables can be incorporated, as well as their mutual correlations. This work takes a different approach and looks into RT-AQF from a machine learning perspective, with historically accurate measurements being used to extrapolate the value of a target variable (particulate matter concentration) both in time (the near future) and space (where there are no sensor measurements). In fact, this approach is particularly useful in such problems due to their chaotic nature: the number of contributing factors is so high that no deterministic model will ever be able to capture all its nuances, and even minor imprecisions on initial conditions will produce vastly different outcomes.

Commonly used methods include classification and regression trees (Burrows et al. 1995), artificial neural networks (Perez and Reyes 2006), fuzzy logic (Shad et al. 2009) and variations of the Kalman Filter (van der Wal and Janssen 2000; Zolghadri and Cazaurang 2006). This paper introduces the use of Gaussian processes as a machine learning tool for air quality forecast, and develops a novel framework for the training and continuous update of the nonparametric model over spatio-temporal data. This framework is able to learn periodical patterns in previous measurements in order to probabilistically predict their values in unobserved areas of the input space, as well as gradually adapt to new behaviors as more information is incorporated into the starting model. A novel training methodology is proposed, named *Structural Cross-Validation*, that is shown to improve accuracy by minimizing an error function while maintaining the spatio-temporal structure of available data. We also explore the use of other environmental variables, such as temperature, humidity and wind, to improve the accuracy of PM10 and PM2.5 predictions.

Overview on Gaussian Processes

The task of estimating air pollutant values from historical data can be seen as a regression problem, where available information is extrapolated to areas of the input space with no direct measurements. While most current approaches use a carefully developed model involving different relevant variables and their correlations, there is a limit to how many variables can be used, and a limit to how complex these correlations can become before the resulting model is rendered computationally unfeasible.

We opted here to go in a different direction and use a nonparametric model, in particular a Gaussian Process (Rasmussen and Williams 2006), to generate the transformation function between inputs and outputs. This approach eliminates the need to explicitly define a model, using instead previous measurements to extract patterns and then extrapolate this information to new, unobserved areas of the input space. This allows for greater flexibility and generality, and we show how to maintain computational complexity under check while achieving a good approximation of the underlying function.

Inference

We start with a set of N observations $\mathbf{y} = \{y_n\}_{n=1}^N$ and its corresponding spatio-temporal coordinates $X = \{\mathbf{x}_n\}_{n=1}^N$, where $\mathbf{x}_n = \{x, y, t\}_n$ contains latitude, longitude and time components. The goal is to estimate a continuous transformation function $f(\mathbf{x}_n)$ between inputs and outputs, thus allowing inference to be performed in unknown regions of the input space. The closer these regions are to the available data, the more accurate the final estimation will be, as it is to be expected.

This proximity is defined by a covariance function $k(x_i, x_j)$, or kernel, which imposes a metric that quantifies how close any two input points are to each other. From this covariance function the covariance matrix K is generated, with components $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$. Inference on a test point $\mathbf{x}_*$ is then of the form:

$$\bar{f}_* = m(\mathbf{x}_*) + K_{*N} \tilde{K}_{NN}^{-1} (\mathbf{y} - m(X)) \qquad (1)$$

$$v_* = K_{**} - K_{*N} \tilde{K}_{NN}^{-1} K_{N*}, \qquad (2)$$

where $\tilde{K}_{NN} = K_{NN} + \sigma_n^2 I$ is the covariance matrix K_{NN} disturbed by zero-mean Gaussian noise of variance σ_n^2 and $m(.)$ is a mean function selected as part of the nonparametric model. The covariance function parameters θ_k, the mean function parameters θ_m and the noise level σ_n, compose the hyperparameter set $\theta = (\theta_k, \theta_m, \sigma_n)$ that defines the GP model. These values can be optimized based on available data by maximizing a certain objective function, usually the log-marginal likelihood (Rasmussen and Williams 2006) due to its balance between complexity and data fit that avoids overfitting.

Covariance Functions

As stated previously, the choice of covariance function plays a pivotal role in the predictive power of a nonparametric model. For the particular case of air pollutant prediction (or any environmental phenomenon), we are particularly interested in temporal trends, that repeat over time in a cyclical manner. For this reason, two different covariance functions were selected, the *Matérn* $3/2$ (Eq. 3) and the *Periodic* (Eq. 4), the first to account for short and medium-term trends and the second to account for cyclical patterns:

$$k_{MT} = \theta_v^2 \left(1 + \frac{\sqrt{3}r}{\theta_l^2}\right) \exp\left(-\frac{\sqrt{3}r}{\theta_l^2}\right), \qquad (3)$$

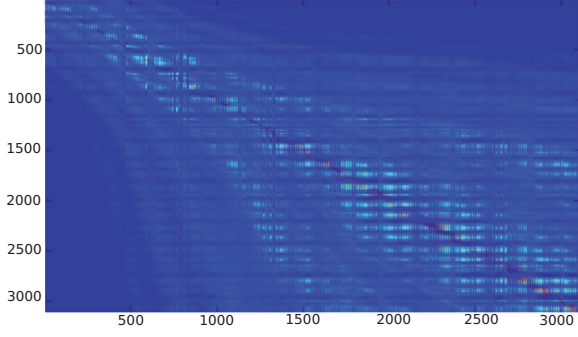$$k_{PR} = \theta_v^2 \exp\left(\frac{-2\sin^2\left(\frac{r}{\theta_p}\right)}{\theta_l^2}\right). \qquad (4)$$

Figure 1: Example of the covariance matrix K, based on Eq. 5. The diagonal area is more closely related, due to the temporal Matérn covariance function, and the various waves indicate different periodic cycles, that gradually lose strength due to decay. The chess pattern is due to spatial correlation, that distantiates points that are physically far away.

In the above equations, $r = |\mathbf{x}_i - \mathbf{x}_j|$ is the distance between input points and the various θ parameters compose the covariance function coefficients θ_k that will be optimized during the training stage. The final spatio-temporal covariance function (Fig. 1) is then of the form:

$$k = \left(k^{xy}_{MT} + k^{xy}_{MT} \right) \times \left(k^t_{MT} + \sum_{p=1}^{P} k^t_{MT} k^{t(p)}_{PR} \right). \quad (5)$$

We can see that the spatial and temporal components of the covariance function are calculated separately and multiplied at the end. Spatially, two Matérns account for short and medium-scale behavioral changes in air pollutant concentration, while temporally there is one Matérn for short-term trends and a series of Periodics to model cyclical patterns. Each Periodic, in turn, is multiplied by a Matérn to account for decay, where the effects of old behaviors slowly fade away as the time interval increases.

Frequency Analysis

Every signal, no matter its complexity, is composed of basic waves, and as more waves are used to represent it, the closer the approximation becomes to the original. However, it is impossible to use all frequencies to describe a signal, and therefore it is necessary to select which ones are more relevant to the application at hand. Also, while the GP framework is capable of learning specific frequency parameters (the θ_p in Eq. 4), it is vital that a good initial estimate is provided due to the presence of local maxima in the objective function.

For this reason, a frequency analysis (Terras 2011) is first conducted on the available data to find its wave components in descending order of relevance (Fig. 2). The first P components are used as starting values for the periodic hyperparameters, and further optimized during the training stage. The upper limit for the number of frequencies used by the nonparametric model is given by the increase in hyperparameter quantity, that affects training times and may create overfitting scenarios.
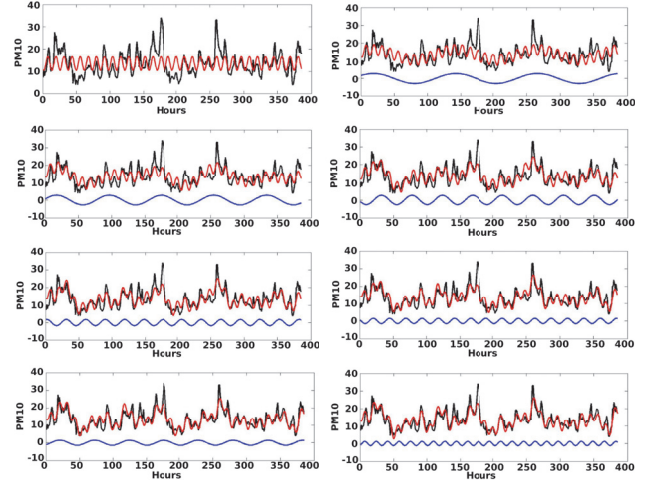

Figure 2: Frequency analysis on hourly measurement of air pollutant concentrations. Starting from the top left, each image incorporates a new frequency into the estimate (blue line), and it is possible to see how the resulting approximation (red line) starts to resemble the true signal (black line).

Sub-sampling

A common drawback in the standard GP derivation (as in most machine learning algorithms) is the computational cost, that scales cubically with the number of training points. This imposes a limit to the amount of available data that can be maintained for training and inference purposes, which is especially detrimental to this application due to the presence of periodic functions that could in theory go back months or even years in time.

For this reason, a novel sub-sampling method is proposed, which exploits the cyclical nature of periodic covariance functions to remove data that will be less correlated to the test points (Fig. 3). This can be achieved by setting up a threshold level δ for the correlation value that is deemed relevant, and discarding all points that fall below this threshold:

$$\{X, \mathbf{y}\}^{t(p)}_{PR} = \left\{ (\mathbf{x}, y)_n \mid k^{t(p)}_{PR}(\mathbf{x}_n, \mathbf{x}_*) > \delta \right\}^{N}_{n=1}. \quad (6)$$

Since the periodic hyperparameters are known (albeit changing during the training stage), they can be used to determine these intervals without the need to calculate every correlation. The resulting subset can then be efficiently obtained by means of simple comparisons between the temporal components:

$$\{X, \mathbf{y}\}^{t(p)}_{PR} = \left\{ (\mathbf{x}, y)_n \mid \bigcup_{i=1}^{M} d(t_* - pi, t_n) < \theta_l h(p) \right\}^{N}_{n=1}, \quad (7)$$

where $d(x, y) = |x - y|$, M is the number of historical periods to be considered and $h(p)$ is the interval of hours considered, which is itself a function of p (empirically we determined $h(p) = \sqrt{p}$ to be a suitable interval function). This process is repeated for each periodic covariance function, generating the subset $\{X, \mathbf{y}\}^{t}_{PR} = \bigcup_{p=1}^{P} \{X, \mathbf{y}\}^{t(p)}_{PR}$ that is
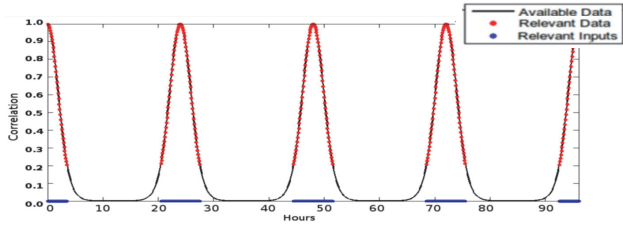
Figure 3: Example of subsampling, for a periodic covariance function with a period $\theta_p = 24$ hours and length-scale $\theta_l = 0.5$. With this configuration, only roughly 30% of the points are used during training and inference.

most correlated to the test points given the necessary computational constraints. This method is particularly useful with longer periods and smaller length-scales, since this configuration tends to create sharp peaks with larger areas of virtually no correlation.

Training Methodology

The aforementioned sub-sampling method, while effective in decreasing the amount of data used for inference, has the drawback of being sensitive to the test point coordinates, since correlation is calculated in relation to that particular timestep. For this reason, we propose a novel training methodology that not only exploits this behavior to produce models optimized for different future predictions, but also maintains during training the same spatio-temporal structure available for inference. We also propose a technique to incorporate extra environmental variables into the prediction of air pollutant concentrations, by combining individually trained base models.

Structural Cross-Validation

The proposed training methodology resembles the traditional Cross-Validation (CV) approach, in the sense that an error function between ground-truth and estimates is minimized. However, this minimization is conducted while maintaining the same spatio-temporal structure that will be available during inference. We start from the design assumption that, during the inference stage, at each timestep k two actions are performed:

1. **Update:** New available measurements are incorporated into the nonparametric model:

$$\{X, \mathbf{y}\}^{1:k} = \left[\{X, \mathbf{y}\}^{1:k-1}, \{X, \mathbf{y}\}^k\right] \quad (8)$$

2. **Prediction:** A new estimate set is obtained for h timesteps into the future (Eqs. 1 and 2):

$$f_*^{k+h} = \mathcal{GP}(\mathbf{x}_*^{k+h}|\{X, \mathbf{y}\}^{1:k}, \theta^k). \quad (9)$$

By constantly adding new information to the nonparametric model, we ensure that predictions will always be at a fixed distance h from available data, and thus will have similar variance values. Gaps in data (such as sensor failure) are treated naturally by an increase in variance values, that

is restored to normal levels once new information becomes available.

For the training stage, each estimate also contains its respective ground-truth, which allows for a third step:

3. **Error Calculation:** The estimate set is compared to their respective ground-truths, according to an error metric $e(.,.)$, and a measure of accuracy ζ is produced:

$$\zeta^k = e(f_*^{k+h}, \mathbf{y}^{k+h}). \quad (10)$$

A diagram of the proposed Structural Cross-Validation algorithm is depicted in Fig. 4. Initially, the available data is divided into two subsets: *starting model* $\{X, \mathbf{y}\}^{1:S}$ and *evaluation data* $\{X, \mathbf{y}\}^{S+1:N}$. Each training step is performed by incrementally (1) incorporating evaluation data from the current timestep into the starting model; (2) performing inference on evaluation data from h timesteps ahead; and (3) comparing these estimates to the actual measurements to generate error values. The accumulated error $\zeta = \sum_{k=S+1}^{N} \zeta^k$ quantifies the adequacy of the current nonparametric model, and is minimized using standard gradient-descend techniques.

Note that, while the final model can be used to perform inference at any spatio-temporal coordinate, it will be optimized for the timestep h used during training. This allows the generation of specific models for different timesteps, that will provide better estimates than a generic one. The same training process can be maintained during the inference stage, taking advantage of the interval between timesteps to further optimize the hyperparameters based on new information. This allows the nonparametric model to gradually adapt to new behaviors and forget old ones, a vital characteristic for long-term usability.

Extra Variables

Here another issue is explored: the incorporation of extra variables into the nonparametric model for the prediction of air pollutant concentrations. Several environmental factors, such as temperature, humidity and wind, may be responsible for changes in air pollutant concentrations, and therefore could be used to improve estimates. A straightforward approach would be the use of multiple-output GPs (Boyle and Frean 2005), but this would create an even bigger computational cost, due to the need of simultaneously maintain all variables and their correlations.
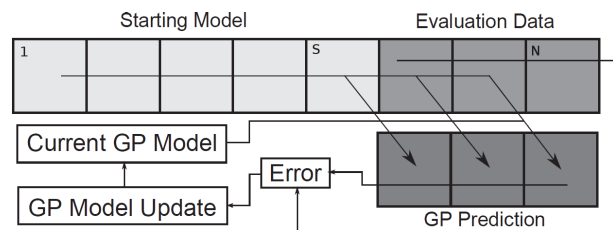

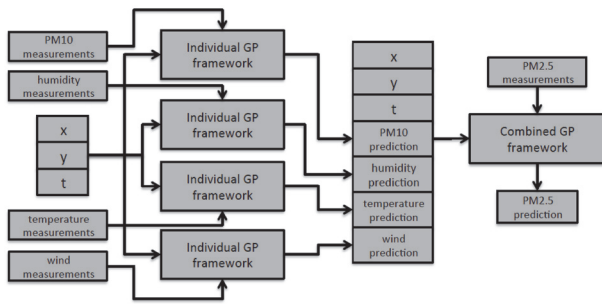Figure 4: Diagram of the Structural Cross-Validation technique.

Figure 5: Diagram of the framework for incorporating extra variables.

Because of that, we propose a layered model that combines different base models to increase the amount of information available for inference, without significantly affecting computational time and memory requirements. A diagram of this layered model is depicted in Fig. 5, for the particular case of using PM10, humidity, temperature and wind to predict PM2.5 concentrations. Each one of these extra variables receives the measurements coordinates $\mathbf{x} = \{x, y, t\}$ and their respective ground-truth to produce its own base model. The predictions from these base models are used to augment the input vector for the combined model, which is then trained based on its own ground-truth information. During inference, further optimization can be performed individually for each base model, and afterwards the new estimates are used to further optimize the combined model.

Experimental Results

Tests were conducted using a sensor network composed of 14 sensors, covering an area of roughly 50×80 km. All sensors provide hourly measurements of PM10, temperature (TP), humidity (HM) and wind (WD), and three of these
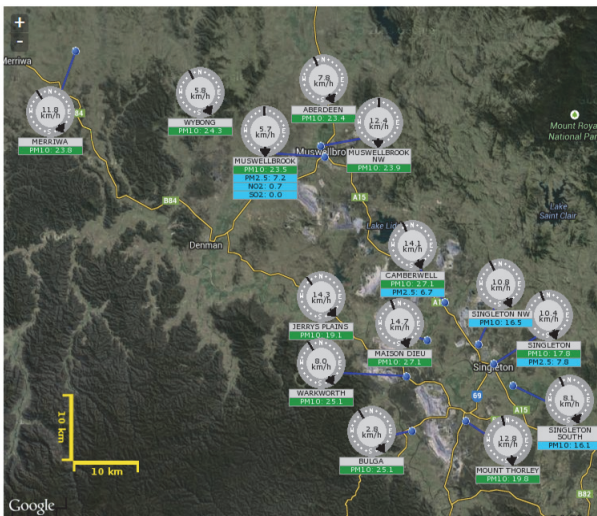

Figure 6: Sensor network used during experiments.

(a) PM10 / 01h

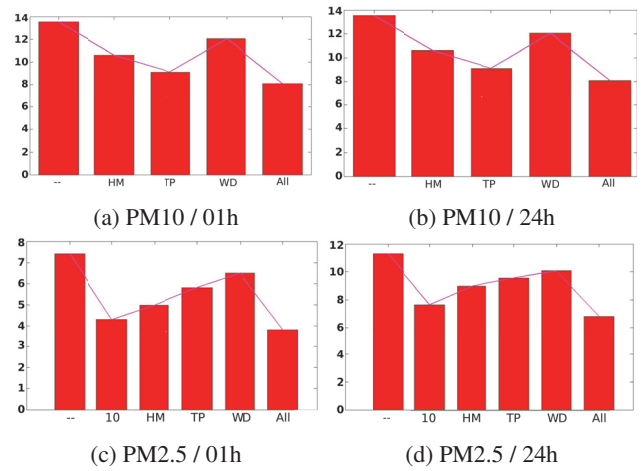(b) PM10 / 24h

(c) PM2.5 / 01h

(d) PM2.5 / 24h

Figure 7: Performance improvement given by the incorporation of extra variables, according to the error metric used during optimization (average over 1 month period).

sensors also provide hourly PM2.5 measurements. The goal was to generate predictive estimates for 1 and 24 hours into the future, as well as in the area surrounding the sensors. In all experiments, a time interval of 3 months was used for training (2 for the starting model and 1 for evaluation), with initial periods of 24, 48, 168 and 672 hours. As new data is incorporated, old data is removed at the same rate. Priority was given to sudden increases in pollutant concentration, since these are the relevant behaviors in air quality control. It is worthy noting that RT-AQF models notoriously struggle with hourly and daily estimates, usually performing better in larger time-scales.

A different base model was trained based on each of these variables, and then combined to generate the final models for PM10 and PM2.5 (PM10 measurements were also used to improve PM2.5 estimates). The improvement generated by the incorporation of each extra variables can be seen in Fig. 7, as well as the improvement generated by the inclusion of all of them. We can see that, for PM10, adding temperature produced the largest improvement, while for PM2.5 it was the introduction of PM10 estimates, followed by humidity. In all cases, the combined model was capable of greatly improving performance, roughly doubling accuracy by maintaining base models of extra variables.

Another way of quantifying the improvement proportioned by the incorporation of extra variables is with QQ-plots; a graphical method for comparing two probability distributions, by plotting their quantiles against each other. If the two distributions are linearly related, the points in the QQ-plot will approximately lie on a line, and any non-linearity will generate deviations. The QQ-plots for base and combined models are shown in Fig. 8, where we can see that the introduction of extra variables is indeed capable of approximating the prediction and measurement distributions. As expected, the algorithm performs less accurately in areas of high pollution value, since these are sporadic and the non-parametric model does not have enough samples for a robust

(a) PM10 / 01h

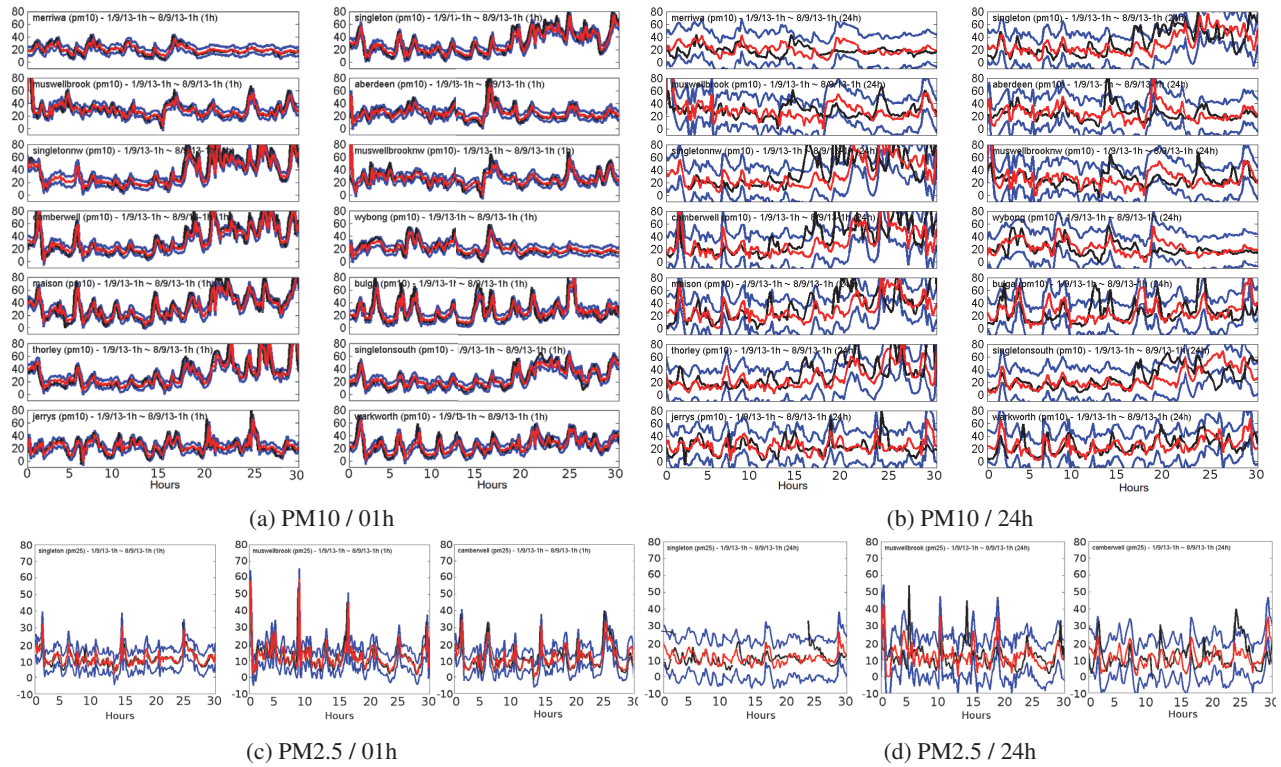(b) PM10 / 24h

(c) PM2.5 / 01h

(d) PM2.5 / 24h

Figure 9: Comparison between GP estimates (red line) and actual sensor measurements (black line), within two standard deviations (blue lines) over the course of 1 month.

estimate. A quantitative comparison between these results and other approaches is given in Table 1.

The progression of predictions over time, for each sensor, can be seen in Fig. 9, alongside the actual measurements and the corresponding 95% confidence boundaries. As expected, 1 hour predictions are very accurate, and this performance decreases as we move away from test data. Note that the estimates are still within the confidence boundaries, indicating

(a) PM10 / 24h (base)

(b) PM10 / 24h (combined)

(c) PM2.5 / 24h (base)

(d) PM2.5 / 24h (combined)

Figure 8: QQ-Plots for PM10 and PM2.5.

Poll. / # hours	SCV-GP	PERS	FITC	CART
PM10 / 1h	**8.48**	14.67	10.42	9.91
PM10 / 24h	**13.21**	22.09	18.29	17.27
PM2.5 / 1h	**3.92**	6.89	5.11	4.83
PM2.5 / 24h	**5.05**	11.93	9.98	9.11

Table 1: Root Mean Square Error comparison between the proposed algorithm (SCV-GP) and the persistent approach (PERS), a sparse GP implementation (FITC, (Quinonero-candela, Ramussen, and Williams 2007)) and decision trees (CART, (Breiman et al. 1984)). The same extra variables were used in all experiments (where applicable).
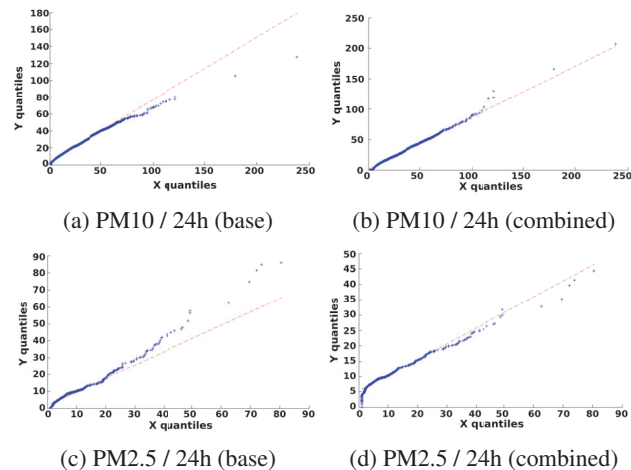
that the nonparametric model is able to correctly gauge the accuracy of its own predictions. These estimates can then be used to prevent high peaks in air pollutant concentrations by taking cautionary measures in areas of highest probability of reaching hazardous thresholds.

Conclusion

We presented a novel algorithm for the online prediction of air particulate matter (PM10 and PM2.5) that relies on a nonparametric model to extract patterns from historical data and uses this information to estimate future behaviors, both spatially and temporally. This is a purely nonparametric approach, without any assumptions regarding the physical and chemical aspects of air pollution, and therefore could be

readily applied to any spatio-temporal predictive scenario. It scales linearly to the number of extra variables, and is able to incorporate new information as it becomes available. The probabilistic nature of GPs allows the generation of uncertainty values alongside the estimates in an intuitive manner that can be easily understood by the final user. In future work we intend to explore the use of mobile sensors, as a way to maximize coverage without the need to increase the network size. More powerful modeling techniques will also be explored, such as nonstationarity and heteroscedasticity.

Acknowledgements

References

Athanasiadis, I. N., and Mitkas, P. A. 2004. An agent-based intelligent environmental monitoring system. *CoRR*.

Athanasiadis, I. N.; Kaburlasos, V. G.; Mitkas, P. A.; and Petridis, V. 2003. Applying machine learning techniques on air quality data for real-time decision support. In *In: First International NAISO Symposium on Information Technologies in Environmental Engineering (ITEE'2003*, 24–27. ICSC-NAISO Publishers.

Boyle, P., and Frean, M. 2005. Multiple output gaussian process regression. Technical report, University of Wellington.

Breiman, L.; Friedman, J.; Olshen, R.; and Stone, C. 1984. *Classification and Regression Trees*. Monterey, CA: Wadsworth and Brooks.

Burrows, W. R.; Benjamin, M.; Beauchamp, S.; Lord, E. R.; McCollor, D.; and Thomson, B. 1995. CART Decision-Tree Statistical Analysis and Prediction of Summer Season Maximum Surface Ozone for the Vancouver, Montreal, and Atlantic Regions of Canada. *Journal of Applied Meteorology* 34:1848–1862.

Cope, M. E., and Hess, G. D. 2005. Air quality forecasting: a review and comparison of the approaches used internationally and in australia. *Clean Air and Environmental Quality* 39:39–45.

Greenbaum, D. S.; Bachmann, J. D.; Krewski, D.; Samet, J. M.; White, R.; and Wyzga, R. E. 2001. Particulate air pollution standards and morbidity and mortality: case study. *Americal Journal Epidemiology* 154(12 Suppl):S78–90.

Marchant, R., and Ramos, F. 2012. Bayesian optimisation for intelligent environmental monitoring. In *IROS*, 2242–2249. IEEE.

McCollister, G., and Wilson, K. 1975. Linear stochastic models for forecasting daily maxima and hourly concentrations of air pollutants. *Atmospheric Environment* 9:417–423.

Perez, P., and Reyes, J. 2006. An integrated neural network model for PM10 forecasting. *Atmospheric Environment* 40(16):2845 – 2851.

Phalen, R., and Phalen, R. 2011. *Introduction to Air Pollution Science*. Jones and Bartlett Learning.

Quinonero-candela, J.; Ramussen, C. E.; and Williams, C. K. I. 2007. Approximation methods for gaussian process regression.

Rasmussen, C. E., and Williams, K. I. 2006. *Gaussian Processes for Machine Learning*. The MIT Press.

Russel, A., and Dennis, R. 2000. Narsto critical review of photochemical models and modeling. *Atmospheric Environment* 34:2283–2324.

Ryan, W. F. 1995. Forecasting ozone episodes in the baltimore metropolitan area. *Atmospheric Environment* 29:2387–2398.

Shad, R.; Mesgari, M. S.; abkar, A.; and Shad, A. 2009. Predicting air pollution using fuzzy genetic linear membership kriging in GIS. *Computers, Environment and Urban Systems* 33(6):472 – 481.

Terras, A. 2011. *Fourier analysis on finite groups and applications*. Cambridge University Press.

U.S. EPA. 2003. Technical assistance document for reporting of daily air quality - air quality index (AQI). Technical report, Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC.

U.S. EPA. 2009. Technical assistance document for reporting of daily air quality - air quality index (AQI). Technical report, Environmental Protection Agency, Office of Air Quality Planning and Standards, Research Triangle Park, NC.

van der Wal, J., and Janssen, L. 2000. Analysis of spatial and temporal variations of PM 10 concentrations in the netherlands using kalman filtering. *Atmospheric Environment* 34(22):3675 – 3687.

Wayland, R. 2002. Communicating real-time and forecasted air quality to the public. *Environment Management*.

Wolff, G. T., and Lioy, P. J. 1978. An empirical model for forecasting maximum daily ozone levels in the northeastern united states. *Journal of Air Pollution Control Association* 28:1034–1038.

Zhang, Y.; Bocquet, M.; Mallet, V.; Seigneur, C.; and Baklanov, A. 2012a. Real-time air quality forecasting, part i: History, techniques, and current status. *Atmospheric Environment* 60:632–655.

Zhang, Y.; Bocquet, M.; Mallet, V.; Seigneur, C.; and Baklanov, A. 2012b. Real-time air quality forecasting, part ii: State of the science, current research needs, and future prospects. *Atmospheric Environment* 60:656–676.

Zhang, Y. 2008. Online coupled meteorology and chemistry models: History, current status, and outlook. *Atmospheric Chemistry and Physics Discussions* 8:1833–1912.

Zolghadri, A., and Cazaurang, F. 2006. Adaptive nonlinear state-space modelling for the prediction of daily mean PM10 concentrations. *Environmental Modelling and Software* 21(6):885 – 894.