Abstract—Increased levels of particulate matter (PM) in the atmosphere have contributed to an increase in mortality and morbidity in communities and are the main contributing factor for respiratory health problems in the population. Currently, PM concentrations are sparsely monitored; for instance, a region of over 2200 km² surrounding Melbourne in Victoria, Australia, is monitored using ten sensor stations. This paper proposes to improve the estimation of PM concentration by complementing the existing high-precision but expensive PM devices with low-cost lower precision PM sensor nodes. Our evaluation reveals that local PM estimation accuracies improve with higher densities of low-precision sensor nodes. Our analysis examines the impact of the precision of the lost-cost sensors on the overall estimation accuracy.

Index Terms—Air pollution, Bayesian maximum entropy, geospatial analysis, kriging, particulate matter, spatiotemporal estimation, wireless sensor networks.

I. INTRODUCTION

According to the World Health Organization, over 4.6 million people die annually as a result of air pollution. High levels of pollutants are a major cause of respiratory conditions such as asthma, bronchitis, and chronic obstructive pulmonary disease [1]–[3]. Concern for pollutants depends on how widespread they are, as well as the legal and regulatory framework of the locale in which they exist. The most widespread pollutants include carbon monoxide, volatile organic compounds, ozone, oxides of nitrogen, sulfur dioxide, and lead. Some primary pollutants can interact to form secondary pollutants, which are produced as a result of the change in chemical balance in the atmosphere. Mixtures of pollutants (smog) can aggravate existing respiratory ailments such as asthma and bronchitis and/or increase the risk of respiratory problems, particularly on “smoggy” days. Elevated lead levels can affect the central nervous system and may impair intellectual development in children [4]. Many studies link air pollutants, such as particulate matter (PM), ozone, nitrogen dioxide, and carbon monoxide (CO), to increases in premature deaths and hospital admissions for people with existing heart and lung disease [4].

One specific form of pollution is PM, which is a fluid mixture of solids and liquids [5], [6]. In particular, PM in the air that has an aerodynamic diameter less than 10 μm, referred to as PM10, and those with a diameter less than 2.5 μm, referred to as PM2.5, are a population health hazard. The PM diameter is smaller than a human hair which averages about 70 μm [6]. PM particles are small enough to penetrate the deepest portions of the lungs, and the very smallest particles can be absorbed into the bloodstream [5]. Exposure to even low levels of PM can cause nasal congestion, sinusitis, throat irritation, coughing, wheezing, shortness of breath, and chest discomfort [6], [7]. Increased mortality and morbidity in communities with elevated PM10 concentrations have been reported by a variety of studies, and PM10 forecasting has been discussed in many articles [8]–[13]. A study conducted by the Harvard School of Public Health in six U.S. cities found that the subjects exposed to higher PM levels were 26% more likely to die prematurely than those exposed to lower concentrations [14]. Therefore, monitoring the PM levels in the environment is an important factor in maintaining good public health.

Regulatory bodies around the world regularly monitor the levels of PM10 concentrations in the atmosphere. While expensive top-end pollution monitoring equipment that are capable of reliably measuring very low levels of pollutants are available, the current practice is that monitoring is done at a low spatial resolution due to the high cost of the equipment. For example, a region of over 2200 km² surrounding Melbourne in Victoria, Australia, is monitored using ten sensor stations (see Fig. 1). The sparsity of these monitoring sites means that pollution levels estimated at locations far removed from the monitoring sites are unreliable, making it difficult to investigate the correlation of PM to population health outcomes. The sparsity of pollution sensors might be augmented by the use of complementary air quality models in both the temporal and spatial domains.

The main objective of this paper is to investigate and propose the use of nodes equipped with low-cost low-precision PM sensors to improve the estimation of PM concentration.
sensors to supplement existing high-precision, often expensive, PM devices with the goal of improving the estimation at finer spatial and temporal resolutions. A more capable sensor network [15]–[29] will enable us to get a more accurate measure of the exposure of individuals to PM10 concentrations and, hence, better understand the impact of PM10 levels on respiratory illnesses. To achieve this objective, we perform simulation studies using synthetically generated PM10 concentration data that match the overall spatiotemporal characteristics of measured PM10 data. In particular, in this paper, we generated synthetic PM10 data that resemble the historical spatiotemporal PM10 data collected by the Environment Protection Authority Victoria (EPA V). We consider PM10 concentration time series from eight sensor stations located in the vicinity of Melbourne. Details of this EPA V sensor deployment can be found at the Environment Protection Authority (EPA) Web site [30]. To generate PM10 over the EPA V monitored region for this study, we propose two space–time (ST) covariance-based noise modeling algorithms: a Fourier transform-based method and a Cholesky decomposition-based method.

Using the generated simulation data, we analyze how the estimation accuracy of PM10 levels at arbitrary locations is affected by the following: 1) spatial resolution of sensors and 2) sensor accuracy. This was done by estimating PM10 levels at a number of arbitrary locations using generated synthetic sensor measurements from high-precision EPA V sensors and a specified number of low-precision sensors placed at locations on a grid. The effect of the number and the accuracy of the low-cost sensors is studied. We used the Bayesian maximum entropy (BME) technique [31], which is a well-accepted technique for the estimation and prediction of spatiotemporally correlated data. Our results show the feasibility of using low-cost low-precision sensors for improving the estimation accuracy of PM10 data.

II. RELATED WORK

Modeling networks of wireless pollutant sensors have recently been a topic of intensive research in applications spanning from environmental and technological monitoring to defense and security systems (see [32] and [33] and references therein). It has been demonstrated that such systems can be described by a mathematical model that has some similarities with population biology. By employing this approach, a number of optimization strategies and operational protocols have been proposed to meet various operational criteria, including the tradeoff between parameters of an individual sensor and the coherent performance of the chemical sensor network and the integrated system (for details, see [32] and [33]). This paper contributes to this line of research by exploring the information gain that can be provided by the networking effect in a system of many sensors, each with low individual accuracy.

A wide variety of sensors and models have been used for the estimation and prediction of air quality and concentrations of specific pollutants [34]–[36]. Slini et al. [8] used a stochastic autoregressive integrated moving average (ARIMA) model to study and make maximum ozone concentration forecasts in Athens, Greece. Diaz-Robles et al. [9] combined an ARIMA model with a neural network (NN) to build a hybrid model to forecast PM10 concentrations in Temuco, Chile. Slini et al. [10] described an operational air quality forecasting module for PM10 based on classification and regression trees and NN methods, which is capable of capturing PM10 concentration trends. Stadlober et al. [11] used multiple linear regression models that combined the information of the present day with meteorological forecasts of the next day to forecast daily PM10 concentrations for sites located in the three cities: Bolzano, Klagenfurt, and Graz, the three capitals in the provinces of South Tyrol, Carinthia, and Styria. Tsai et al. [12] studied PM2.5 and PM10 in Bangkok using simple regression and correlation analysis for data gathered in a shopping center and at a university. Yu et al. [37] studied the spatiotemporal distribution of PM2.5 across the Taipei area from 2005 to 2007 using a land use regression model. They established a quantitative relationship between PM2.5/PM10 and land use information for predicting the PM2.5 concentration. Epitropou et al. [38] used a fusion strategy to select environmental information from open access web resources of various types and data from monitoring stations in order to provide tailored information to the users. Forecast models are produced using published air quality information in the form of color-mapped georeferenced images and data from chemical weather databases.

Multivariate geostatistics offers many alternatives for choosing the family of functions $F$ and a best member $F^* \in F$ to describe and model the observations. Good introductions to some of these methods are the texts by Clark [39] and Wackernagel [40]. The most popular and perhaps best known possibilities are variograms, ordinary, simple, and weighted kriging, principal component analysis, canonical analysis, and correspondence analysis. Christakos et al. [31], [41] present a method for estimating PM10 levels in the state of California based on an integrated ST domain using the BME mapping approach. Their BME method is a generalized model, in which kriging becomes one of the special cases. We use this method to estimate the concentrations at unknown spatiotemporal locations in the EPA V region.

One of the primary challenges in many of these studies (including ours) is that analysis is based on measurements from a few sparsely located pollution monitors. For example, one such study in Santiago, Chile, used the average of just five
monitoring sites to obtain pollution data representing the whole city [42]. Localized variations of pollutant concentrations may not be adequately captured by a small number of sparsely located monitoring sites; therefore, in [42], Santiago is divided into five regions, each centered around a pollution monitoring station. This strategy seemed to improve correlations between pollutant levels and observed health effects throughout the city. Our approach is somewhat different. We first validate our model with simulated data at the nodes of a regularized rectangular grid that circumscribes the sensor stations shown in Fig. 1. Then, we use the time series at the EPA V sites as a basis for both interpolation and extrapolation over the spatial domain of Fig. 1.

The statistical properties of PM (pollutant, aerosol, moisture, and dust) in the atmospheric boundary layer have been investigated; the pioneering work of Richardson and Batchelor [43], [44] is well documented. It is recognized that these properties and statistics of associated particle time series emerge from the physics of turbulent mixing and particle dispersion by random flow [45], [46]. This enables the development of physics-based frameworks for data analysis and implementation of rigorous computer simulations for the concentration time series (see [45], [47], [48], and references therein). In the context of the current study, two specific properties of particle statistics are most relevant: 1) The concentration time series are strictly positive (since measured values are simply counts of particles), and 2) they can be comprehensively characterized by the first two statistical moments [46] (since they are driven by the underlying Gamma distribution). The latter property becomes critical for the application of the BME model; see (7).

III. PROBLEM STATEMENT AND APPROACH

A. Problem Statement

We consider a ST domain where a general point in space and time is given by \( \mathbf{p} = (s, t) \in S \times T \), where \( s = (x, y) \) is the geographical location and \( t \) is the time. A set of \( N_h \) high-precision sensors and \( N_l \) low-precision sensors is deployed at known geographical locations in the region, and measure PM10 concentrations over \( T \) discrete time points. Hence, a collection of \( n_h = N_h \times T \) high-precision sensor measurements and \( n_l = N_l \times T \) low-precision sensor measurements is collected at \( n_d = n_h + n_l \) ST points. The aim is to estimate the PM10 concentrations at an arbitrary set of \( n_k \) ST points, called the estimation points, based on the measured values from \( n_d \) ST points. Note that \( n_k = N_k \times T \), where \( N_k \) is the number of geographic (estimation) locations.

We denote the ST random field realizations (PM10 concentrations) over the ST region as a collection of correlated random variables \( Z_{\text{map}} = \{ Z_{\text{data}} \cup Z_{\text{est}} \} \) at points \( \mathbf{p}_{\text{map}} = \{ \mathbf{p}_{\text{data}} \cup \mathbf{p}_{\text{est}} \} \), where \( Z_{\text{data}} = \{ Z_i, Z_{i+1}, \ldots, Z_{i+N_d} \} \), \( Z_{\text{est}} = \{ Z_{n_d+1}, \ldots, Z_{n_d+k}, \ldots, Z_n \} \), \( \mathbf{p}_{\text{data}} = \{ p_{1}, p_{2}, \ldots, p_{n_d} \} \), and \( \mathbf{p}_{\text{est}} = \{ p_{n_d+1}, \ldots, p_{n_d+k}, \ldots, p_{n_k} \} \).

We define the root-mean-square estimation error \( \epsilon_s \) at a spatial location \( s \) over a time period \( T \) as

\[
\epsilon_s = \sqrt{\frac{1}{T} \sum_{i=1}^{T} (Z(s, t_i) - \hat{Z}(s, t_i))^2}
\]

where \( Z(s, t_i) \) is the true value of the PM concentration at an ST point \( (s, t_i) \) and \( \hat{Z}(s, t_i) \) is the estimated value of the PM concentration.

The overall root-mean-square error (over space and time) \( \epsilon \) for a given sensor configuration is defined as

\[
\epsilon = \sqrt{\frac{1}{N_k T} \sum_{i=1}^{N_k} \sum_{i=1}^{T} (Z(s_i, t_i) - \hat{Z}(s_i, t_i))^2}.
\]

Our aim is to study the impact of the number of low-precision sensors \( (N_l) \) and their accuracy on the estimation errors \( (\epsilon, \epsilon_s) \). In particular, we are interested in determining whether supplementing high-precision sensors with low-precision sensors can reduce the error.

B. Our Approach

To study the impact of low-precision sensors on estimation accuracy, we need to compare estimated PM values to actual concentrations (ground-truth data) at chosen locations. While the measurements from high-precision sensors are the best possible estimates of actual concentrations, due to the spatial sparseness of these measurements, there are not enough ground-truth data available to perform the intended analysis. Therefore, the approach that we have taken is to generate synthetic ground-truth data that resemble the characteristics of real data. In this paper, we use the real PM10 data available from eight sensors (later, we will identify the eight sensors that we are using that have significantly less missing data) deployed in the region around Melbourne, Australia, by the EPAV. We describe the techniques that we use for this purpose in Section IV.

After generating synthetic ground-truth data, the simulated concentration time series corresponding to \( N_h \) EPAV sensor station locations are taken as the measurements from the high-precision sensors. We then apply a sensor model to simulate low-precision sensor measurements made at \( N_l \) additional locations. The sensor model allows us to mimic characteristics such as noise, saturation, linearity, and the minimum measurement threshold of sensors. Then, we use simulated measurements corresponding to both high- and low-precision sensors to estimate PM10 concentrations at a set of arbitrary known locations not coinciding with the existing sensor locations. We use the BME technique to estimate the concentrations at chosen locations. The BME technique is described in Section VI.

IV. SYNTHETIC GROUND-TRUTH DATA GENERATION

We propose two ST covariance-based noise modeling algorithms to generate ground-truth data that simulate PM10 concentrations over the EPAV monitored region. One is based on the fast Fourier transform (FFT) model, and the other is based on the Cholesky factorization model. We now formulate the two ST covariance-modeled noise schemes.

A. FFT Model

Assume a wide sense stationary (WSS) 2-D covariance model \( B(\Delta x, \Delta y) \) for the spatial region. If we assume that
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

B is also isotropic, then it can be written as \( B(r) \), where \( r = \| x - y \|_2 \) is the Euclidean or 2 norm of the vector \( (x, y) \). One such model could be the Gaussian \( B(r) = ve^{-r^2/a^2} \), where \( v \) and \( a \) are the parameters of the model.

The most popular way to simulate WSS covariance-modeled noise is by using the FFT. Assume a grid of sensors \( S = \{s_{11}, s_{12}, \ldots, s_{n(m-1)}, s_{nm}\} \), where \( s_{ij} = (x_i, y_j) \) is the sensor location. First, we construct the \( n \times m \) matrix \( B = [B(r_{ij})]_{n \times m} \), where \( r_{ij} = \| s_{ij} - s_{i,j} \|_2 \) and \( s_{ij} \) is the center of the grid. Then, the covariance-modeled noise is constructed as follows:

1. Compute the 2-D FFT of \( B, \xi_B = \text{FFT}_2(B) \).
2. Compute the 2-D FFT of an \( n \times m \) complex Gaussian white noise matrix \( W, \xi_W = \text{FFT}_2(W) \).
3. Compute the 2-D inverse FFT, \( V = I\text{FFT}_2(\sqrt{\xi_B} \odot \xi_W) \), where \( \odot \) is the Hadamard product.

This produces two noise draws: the real part and the imaginary part of \( V \) (here, we use the real part of \( V \)).

To add time dependence to this model, we simply construct an \( n \times m \times T \) matrix \( B = [r_{ijw}]_{n \times m \times T} \), where, now, \( r_{ijw} = \| s_{ijw} - s_{i,j}\|_2 \) and \( s_{ijw} = (x_i, y_j, t_w) \). The computation process is the same as above, except that 3-D FFTs and inverse FFTs are used, and a 3-D complex Gaussian white noise random matrix is drawn in step 2.

Fig. 2 shows examples of FFT-simulated time-series data on a grid. The advantage of the FFT-based model is that it is computationally fast and easy to implement. However, the FFT-based model is limited to a grid arrangement, which may not be appropriate for simulations of real systems. Hence, we now move to a different type of model (that produces equivalent results for arbitrary sensor network locations).

**B. Cholesky Factorization Model**

Now, consider some collection of sensors \( S = \{s_1, \ldots, s_n\} \), where \( s_i = (x_i, y_i) \) is the sensor location (not necessarily grid based). First, we construct an \( n \times n \) matrix, \( B = [B(r_{ij})]_{n \times n} \), where \( r_{ij} = \| s_i - s_j \|_2 \). \( B \) is positive semidefinite and thus can be factorized into \( B = RR^T \) (Cholesky factorization). Draw an \( n \)-length vector \( \beta \) of unit-variance zero-mean normally-distributed random numbers, and compute \( \phi = R \beta \). The covariance of \( \phi \) is

\[
B_\phi = \langle \phi \phi^T \rangle = R(\beta \beta^T) R^T = R I_n R^T = B
\]

where \( \langle \cdot \rangle \) indicates the expected value. Hence, the covariance of \( \phi \) is exactly \( B \). The vector \( \phi \) is a set of simulated sensor measurements at each location in \( X \) according to the statistical model \( B(r) \).

Now, assume a time series of sensor measurements \( S = \{S_1, \ldots, S_T\} \), where \( S_t \) is the \( t \)-th set of measurements. Consider the time-based covariance model \( B_t(\Delta t) \). Construct the matrix \( B_t = [B_t(\Delta t_{ij})]_{T \times T} \), where \( \Delta t_{ij} = |t_i - t_j| \), factor it as \( B_t = R_t R_t^T \), and make a random draw \( \Phi_t = R_t \beta \), where \( \beta \) is a \( T \times n \) matrix of unit-variance zero-mean normally distributed random numbers. Each row of \( \Phi_t \) is an \( n \)-length vector of \( B_t(0) \)-variance zero-mean normally-distributed random numbers. Therefore, \( \langle \Phi_t^T(\Phi_t)_t \rangle = B_t(0) I_n, \forall t \), where \( \Phi_t \) is the \( t \)-th row of \( \Phi \). Finally, we can create ST covariance-modeled noise by \( \Phi = R \Phi_T \), where \( \Phi = \{\phi_1, \ldots, \phi_T\} \) is the time series of simulated sensor measurements. Each \( \phi_t \) has the spatial statistics of \( B(r) \), and the time series at each sensor location has the temporal statistics of \( B_t(\Delta t) \). The advantage of this method is that it is not constrained to a grid-based sensor arrangement (as the FFT model is). Fig. 3 illustrates using the Cholesky method to simulate the sensor measurements of the EPA Victoria sensor array (ten locations).

**V. MODEL FITTING TO REAL EPA DATA**

The first step to fitting a WSS spatially isotropic spatiotemporal covariance model to real data is to compute the variogram (covariance estimate) of the sensor measurements \( z(s, t) \). Since we are assuming an isotropic WSS model of the form \( B(r, t) = B_s(r)B_t(\Delta t) \), we can model the spatial and temporal covariances separately. The basic steps of this process are as follows.

1. Compute the spatial variogram \( \hat{B}_s(r) \) and temporal variogram \( \hat{B}_t(\Delta t) \) from the data \( z(s, t) \).
2. (Optional) Interpolate the variogram(s) onto a uniform grid.
3. Regress parameterized models to each of \( \hat{B}_s(r) \) and \( \hat{B}_t(\Delta t) \).
4. Choose the best spatial model and the best temporal model (if more than one model were regressed to each variogram).

In detail, the first step is to calculate the variograms, viz., covariance estimates. The isotropic spatial covariance estimate
Step 2 of the model-fitting process and 48 h. This is intuitively pleasing as this corresponds to the periodic nature, indicating strong correlation peaks at time lags of 12, 23, 36, and 48 h. Interestingly, the temporal variogram exhibits a periodic nature, with variance estimates at integer-valued time lags settled to near 0 at the next lag point of 5.

The temporal covariance estimate is computed as

\[ \hat{B}_t(\Delta t) = \sum_{m=1}^{n_s} \sum_{\forall i,j} \frac{(z(s_m, t_i) - \mu_z)(z(s_m, t_j) - \mu_z)}{n_s n \Delta t} \]  

(5)

where \( n_r \) is the number of summed terms in the second summation (or the number of covariance terms averaged for the lag \( r = \sqrt{||s_i - s_j||^2} \) and \( \mu_z \) is the mean of the sensor measurements. The temporal covariance estimate is computed as

\[ \hat{B}_s(r) = \sum_{t=1}^{T} \sum_{\forall i,j} \frac{(z(s_i, t) - \mu_z)(z(s_j, t) - \mu_z)}{T n_r} \]  

(4)

where \( n_r \) is the number of summed terms in the second summation for the lag \( r = (t_i - t_j) \).

Fig. 4(a) and (b) shows the spatial and temporal covariances for the EPA V data (described in Section VII), estimated over 100 consecutive time slices. What we see right away in view (a) is that there are very few data points in the spatial lag region of \( 0 < r < 16.5 \). Also, the spatial covariance estimate shows that the correlation between sensor measurements has already settled to near 0 at the next lag point of \( r = 16.2 \). This imposed a difficult modeling task because the behavior of the PM10 concentrations between \( r = 0 \) and \( r = 16.5 \) is unknown (this is exactly the reason that a dense sensor network is needed). In contrast to the sparse spatial estimate, the temporal covariance estimate is relatively dense (obviously, we have temporal covariance estimates at integer-valued time lags \( \Delta t = 0, 1, 2, \ldots \)). Interestingly, the temporal variogram exhibits a periodic nature, indicating strong correlation peaks at time lags of 12, 23, 36, and 48 h. This is intuitively pleasing as this corresponds to the periodic nature of the day. Step 2 of the model-fitting process is used to alleviate the difficulty of fitting a model to the spatial covariance estimate shown in Fig. 4(a). Because there are no data points in \( 0 < r < 16.5 \), we interpolate \( \hat{B}_s(r) \) onto a uniform grid \( r = 0 \) to 16.5 as shown in Fig. 4(c). The interpolated data (using cubic interpolation) give the model-fitting algorithm data to work with the region \( 0 < r < 16.5 \). Is this really what the real-world covariance would look like? We do not know, but this at least gives us a reasonable estimate. In step 3, we regress a set of predetermined models to each of \( \hat{B}_s(r) \) and \( \hat{B}_t(\Delta t) \). The models that we use are shown in Table I. To fit a model to the covariance estimate \( \hat{B}(q) \)—i.e., either \( \hat{B}_s(r) \) or \( \hat{B}_t(\Delta t) \)—we seek to minimize the error function

\[ E_m(\Pi) = \sum_q n_q^2(\hat{B}(q) - B_m(q, \Pi))^2 \]  

(6)

where \( q \) represents all possible lags, \( r \) or \( \Delta t \), and \( B_m(q, \Pi) \) is one of the seven models shown in Table I. In essence, we are weighting each summation term in \( E_m \) by the number of pairwise estimates used to compute \( \hat{B}(q) \) at lag \( q \). Recall that we first interpolated \( \hat{B}_s(r) \) onto a uniform grid before fitting to the models. For the spatial covariance, we thus set \( n_r = 1 \), \( \forall r \).

Table I

<table>
<thead>
<tr>
<th>Name</th>
<th>Parameters</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exponential</td>
<td>( (v, a) )</td>
<td>( B_1(q, \Pi) = ve^{-q/a} )</td>
</tr>
<tr>
<td>Gaussian</td>
<td>( (v, a) )</td>
<td>( B_2(q, \Pi) = ve^{-q^2/a^2} )</td>
</tr>
<tr>
<td>Cosine Hole</td>
<td>( (v, a) )</td>
<td>( B_3(q, \Pi) = v \cos(\pi q/a) )</td>
</tr>
<tr>
<td>Sinc</td>
<td>( (v, a) )</td>
<td>( B_4(q, \Pi) = v \sin(\pi q/a) / (\pi q/a) )</td>
</tr>
<tr>
<td>Nugget</td>
<td>( (v) )</td>
<td>( B_5(q, \Pi) = v, B(q) = 0, q &gt; 0 )</td>
</tr>
<tr>
<td>Spherical</td>
<td>( (v, a) )</td>
<td>( B_6(q, \Pi) = v(1 - 3q/2a - 1/2q/a^2)^3 )</td>
</tr>
<tr>
<td>Mexican Hat</td>
<td>( (v, a) )</td>
<td>( B_7(q, \Pi) = v(1 - aq^2)e^{q^2/a^2} )</td>
</tr>
</tbody>
</table>

Note that, in (4), every \( \hat{B}(q,r) \) term in the second summation for the lag \( r = 0 \) to 16.5 is used to alleviate the difficulty of fitting a model to the spatial covariance estimate shown in Fig. 4(a). Because there are no data points in \( 0 < r < 16.5 \), we interpolate \( \hat{B}_s(r) \) onto a uniform grid \( r = 0 \) to 16.5 as shown in Fig. 4(c). The interpolated data (using cubic interpolation) give the model-fitting algorithm data to work with the region \( 0 < r < 16.5 \). Is this really what the real-world covariance would look like? We do not know, but this at least gives us a reasonable estimate. In step 3, we regress a set of predetermined models to each of \( \hat{B}_s(r) \) and \( \hat{B}_t(\Delta t) \). The models that we use are shown in Table I. To fit a model to the covariance estimate \( \hat{B}(q) \)—i.e., either \( \hat{B}_s(r) \) or \( \hat{B}_t(\Delta t) \)—we seek to minimize the error function

\[ E_m(\Pi) = \sum_q n_q^2(\hat{B}(q) - B_m(q, \Pi))^2 \]  

(6)

where \( q \) represents all possible lags, \( r \) or \( \Delta t \), and \( B_m(q, \Pi) \) is one of the seven models shown in Table I. In essence, we are weighting each summation term in \( E_m \) by the number of pairwise estimates used to compute \( \hat{B}(q) \) at lag \( q \). Recall that we first interpolated \( \hat{B}_s(r) \) onto a uniform grid before fitting to the models. For the spatial covariance, we thus set \( n_r = 1 \), \( \forall r \).
as each $n_{\Delta t}$ is different, with $n_0 = T_1$, $n_1 = T - 1$, $n_2 = T - 2$, etc. The minimization of the function in (6) is accomplished using a quasi-Newton line search method.

After we regress each model to each of $B_s(r)$ and $B_t(\Delta t)$, we choose the best model for each covariance estimate. Let $\Pi^*_m$ be the best parameters found by minimizing (6) for either the spatial or temporal covariance. Then, the best overall models are chosen by $m^* = \arg \min_m E_m(\Pi^*_m)$, which is solved for each of the spatial and temporal covariances. Fig. 5(a) and (b) shows the best empirically determined model for each of $B_s(r)$ and $B_t(\Delta t)$. The chosen model for the spatial covariance is the exponential model $B_1(r, (1, 12))$, and the chosen model for the temporal covariance is the exponential model $B_1(\Delta t, (1, 2.4))$. Note that, in both cases, $\nu = 1$. This is because we normalize the data to unit variance. For the estimates shown in Fig. 4, $B_s(0) = B_t(0) = 147$.

As Figs. 4(c) and 5(a) show, the interpolation method most likely overestimates the parameter $a$ of the exponential spatial covariance model. However, given the absence of covariance estimates at spatial lags $0 < r < 16.5$, we believe that any value $5 \leq a \leq 12$ would provide a reasonable fit for these data. Fig. 5(b) shows that the exponential model fails to capture the periodic nature of the temporal covariance model. In the future, we plan to extend our modeling and simulation capabilities to more accurately match this periodic behavior.

Fig. 5(c) and (d) shows the spatial and temporal covariance models compared to the sample-based covariance estimates of a random field generated by the FFT process described earlier. We generated 51 time slices of a 101 by 101 grid, based on the model parameters estimated from the EPAV data [shown in Fig. 4(a)]. Then, the sample-based covariance was estimated from the simulated field. The figure shows that the covariance of our simulated random field matches well with the model.

VI. PM CONCENTRATION ESTIMATION

In order to estimate the PM concentration at arbitrary locations, we use a recently developed approach called BME, which involves the concepts of Bayesian conditionalization and entropy [31], [49]. The definitive works on this are those by Serre and Christakos [31], [49], who have used this approach quite successfully for PM10 estimation in various locales [41].

The BME approach incorporates two prime knowledge bases (a collection of knowledge sources): 1) a general knowledge base $G$, such as that from physical laws, summary statistics, and scientific theories, and 2) specificatory knowledge or case-specific knowledge base $\ominus$, such as that obtained through experience with specific situations. The total knowledge base is $H = G \cup \ominus$. Furthermore, the specific knowledge base $\ominus$ accounts for both hard data (measured values that we are willing to concede are correct with near certainty) and soft data (interval or probabilistic data that are uncertain due to, for example, inexact knowledge, experience, intuition, low-precision sensor outputs, etc.) [49].

Let the ST random field realizations be represented as a collection of correlated random variables $Z_{\text{map}} = (Z_1, Z_2, \ldots, Z_n, Z_k)$ at points $p_{\text{map}} = (p_1, p_2, \ldots, p_n, p_k)$. The aim of BME is to obtain a spatiotemporal map displaying the estimate at a point $p_k \in p_{\text{map}}$ of the unknown value $Z_k$ based on the observed values $Z_1, Z_2, \ldots, Z_n$, at locations $p_1, p_2, \ldots, p_n$. For each point in $p_{\text{map}}$, the ST model assigns a probability value using the probability density function (pdf) $f_{\hat{H}}(Z_{\text{map}})$.

There are three main stages of knowledge acquisition, interpretation, and processing involved in the BME approach: (1) the prior stage; (2) the meta-prior stage; and (3) the integration or posterior stage.

In the prior stage, ST analysis starts with the basic set of assumptions and the general knowledge base $G$. The expected information contained in the map $Z_{\text{map}}$ is expressed as $\mathcal{E}(f_G[Z_{\text{map}}]) = -\int f_G(Z_{\text{map}}) \log f_G(Z_{\text{map}}) dZ_{\text{map}}$, where $f_G$ is the multivariate pdf model associated with $G$ before the specific knowledge base $\ominus$ has been taken into consideration and $\mathcal{E}(\cdot)$ denotes the expectation operator. The shape of the prior pdf $f_G$ is determined by maximizing the expected prior information subject to the physical constraints $\mathcal{E}(g_\alpha(Z_{\text{map}})) = \int g_\alpha(Z_{\text{map}}) f_G(Z_{\text{map}}) dZ_{\text{map}}$, where $g_\alpha$ is a set of known functions of $Z_{\text{map}}$ with $g_0 = 1$, $\mathcal{E}(g_0(Z_{\text{map}})) = 1$, and $\alpha = 0, 1, 2, \ldots, N_c$. The total number of $N_c$ is such that stochastic moments are included that involve all the ST points $p_k \in p_{\text{map}}$ of the map. In this paper, the $G$ consists of the mean functions $m_\theta(p)$ and the covariance functions $C_\theta(p, p')$ given by

$$C_\theta(p, p') = \mathcal{E}[(Z(p) - m(p))(Z(p') - m(p'))]$$

where $m(p) = \mathcal{E}[Z(p)], \forall p$. Hence, the $g_\alpha$ is chosen such that the expectation $\mathcal{E}(g_\alpha(Z_{\text{map}}))$ provides the ST mean and covariance functions throughout the ST domain of interest. The optimization is performed using the method of Lagrange multipliers, and then, the expression for the prior pdf becomes $f_G(Z_{\text{map}}) = J^{-1} e^{-\nu_0} \gamma_G[Z_{\text{map}}]$, where $J = e^{-\mu_0}$, $\gamma_G[Z_{\text{map}}] = \sum_{n=1}^{N_c} \mu_n g_n(Z_{\text{map}})$, and $\mu_n$ are the Lagrange multipliers.

In the meta-prior stage, the specificatory knowledge $\ominus$ is considered. The pdf $f_G$ is updated by means of Bayesian conditionalization. In the integration or posterior stage, both of these knowledge are incorporated to obtain the...
posterior pdf \( f_H(Z_{\text{map}}) = Q^{-1} Y_\emptyset | Y_\emptyset, \emptyset, Z_{\text{map}} \), where \( Q \) is a normalization parameter. If we assume that the specific knowledge \( \emptyset \) consists of the hard data and the probabilistic soft data, then the posterior operator \( Y_\emptyset \) becomes \( Y_\emptyset = Q^{-1} \int_I e^{Y_\emptyset | Z_{\text{map}}^d} f_{z_\emptyset} | Z_{\text{soft}}^d \), where \( \emptyset \) is a hard data set, \( I \) denotes the domain of \( Z_{\text{soft}}^d \), \( Z_{\text{hard}} \) denotes the hard data, \( Z_{\text{data}} = \{ Z_{\text{hard}} \cup Z_{\text{soft}} \} \subset Z_{\text{map}}, \) and \( F_{z_\emptyset} \) is the cumulative distribution function obtained from \( \emptyset \). Several possibilities exist to obtain the posterior estimate \( \hat{Z}_k \) at an unknown point \( p_k \) using \( f_H(Z_{\text{map}}) \). One such possibility is to use the mean estimate at the point \( p_k \), which is given by \( \hat{Z}_{k,\text{mean}} = \int Z_k f_{H}(Z_k) dZ_k \). In this paper, this mean estimate is used to find the PM10 concentrations at the off-sensor locations. Note that this BME mean estimate minimizes the mean squared estimation error.

Let the ST variability of the PM10 concentration \( Z \) be described in terms of a centered covariance function \( C(\kappa, \tau) = E[(Z(p) - \mu(p))(Z(p') - \mu(p'))] \), where \( \mu(p) = (s - s'), (t - t') \) is the mean. The quantities \( \kappa \) and \( \tau \) are space and time lags and show that this covariance is spatially isotropic and stationary in time. For the case of \( \kappa = 0 \), \( C(0, \tau) \) is reduced to the single-point time series (observations from an individual sensor) and can easily be validated experimentally [45]. More specifically, the difference \( C(\kappa, \tau) - C(0, 0) \) should exhibit a linear dependence on its arguments at the limit of small values of \( \kappa, \tau \) [46]. Since this property has been reliably recovered experimentally [45], it can be used as a supporting argument for the proposed model.

We use \( C(\kappa, \tau) = v_s \gamma \alpha_s e^{-r/a_s} e^{-\tau/a_t} \) to model the covariance structure in the evaluation. The covariance coefficients \( v_s \) and \( v_t \) provide information about variations in the PM10 distribution. The coefficient \( \alpha_s \) is the range of spatial variation, and \( \alpha_t \) represents temporal fluctuations. We match the coefficients to our generated simulation data, which are a representation of the EPAV PM10 data (as explained in Section V), and utilize the BME software for the estimation [50].

VII. EVALUATION

A. Representation of the EPAV Sensor Network

Our work is facilitated by introducing the artificial Cartesian coordinate system for the EPAV nodes (shown in Fig. 1) on a regular grid of size 200 \( \times \) 200. The EPAV nodes considered in this evaluation are deployed at the following eight locations (including the coordinates): Alphington (141, 171), Box Hill (159, 156), Brighton (135, 122), Dandenong (177, 89), Mooroolbark (197, 170), Deer Park (89, 176), Footscray (115, 164), and Richmond (135, 156). The EPAV PM10 data (high-precision sensor data) were collected during the period between October 3, 2011 and April 20, 2012 at a sampling interval of 1 h. We use the FFT-based model (grid sensor-based scheme; see Section IV-A) for generating the ground-truth data for our experiments as the given sensor locations are already on a Cartesian grid.

The ground-truth data are generated for a 3-D rectangular parallelepiped (rectangular box) region spanned by the coordinates \([1, 200] \times [1, 200] \times [1, 50]\), where the \( x \)- and \( y \)-axis ranges are \([1, 200]\) each and the third axis (time slices) range is \([1, 50]\). The grid spacing is 1 in all three directions. The spatial covariance model \( v_s \gamma \alpha_s e^{-r/a_s} \) has two parameters with values \( v_s = 1 \) and \( \alpha_s = 12 \), and the temporal covariance model \( v_t e^{-\tau/a_t} \) has parameter values \( v_t = 1 \) and \( a_t = 2.4 \). These parameters were selected based on the model fitted to the spatiotemporal variogram curves obtained from the real EPAV PM10 data (see Section V).

B. Representation of High- and Low-Precision Sensors

In order to study the impact of additional low-precision sensors on the estimation accuracy, we consider a deployment which has a combination of high-precision sensors as well as low-precision sensors. The high-precision sensors are located at the actual EPAV sensor locations, and their concentration values are set to the sensor concentrations of the generated data (i.e., consider them as ideal sensors with zero error). In the BME estimation, these are the hard data locations. To study the impact of low-precision sensors, we place them on a regular grid encompassing the EPAV sensor network. The number of low-precision sensors \( N_l \) is varied by changing the grid size. The precision of the sensors is controlled by adding zero-mean Gaussian noise with varying standard deviation. For BME estimation, these sensors are the soft data.

C. Effect of the Number of Low-Precision Sensors \( N_l \)

In this experiment, the estimation locations are a set of grid locations with a spatial grid spacing of 20 within the spatial grid of size \([80, 200] \times [80, 200]\). The soft data locations are the grid locations with the number of horizontal and vertical spacings selected from a range of values between 1 and 100, i.e., we use square grids of size \( 1 \times 1 \) to \( 100 \times 100 \) for the low-precision sensors. If a soft data location coincides with a hard data location, the soft data sensor is removed. The soft data are modeled using a Gaussian pdf with the generated ground-truth value as the mean and a constant value of 0.2 as the standard deviation (\( \gamma \)). The estimated values are compared with the generated ground-truth values obtained at these estimation locations. The root mean square error (RMSE) \( \epsilon \) [see (2)] is computed across all the estimation locations that are noncoinciding with the soft or hard sensor locations and all the time instances, for each grid size of the soft sensors (number
of soft data locations). The normalized error for each grid size of the soft sensors is computed by dividing the RMSE $\epsilon$ by the RMSE obtained for the hard only case, where there are no low-precision sensors deployed, i.e., $N_I = 0$. Fig. 6 shows the normalized error versus the number of soft sensors. As intuition would suggest, the normalized error decreases as the number of soft sensors is increased.

The intensity plots in Fig. 7 show the root-mean-square estimation error $\epsilon_s$, at spatial locations for the cases $N_I = 0$ (no low-precision sensors), $N_I = 8 \times 8$, $N_I = 15 \times 15$, and $N_I = 25 \times 25$. The estimation locations used are a $50 \times 50$ grid.

**D. Effect of Sensor Noise on the Estimation**

In this experiment, the estimation locations are a set of grid locations with a spatial grid spacing of 20 within the spatial grid of size $[80, 200] \times [80, 200]$. The soft data are modeled using a Gaussian pdf whose mean is the generated ground-truth value (using the FFT algorithm). The sensor noise, standard deviation $\gamma$, is changed in the range $[0.001, 1]$ at intervals of 0.01. Five different soft sensor configurations are considered. Each configuration consists of soft sensor locations at the sets of equally spaced grid locations with a total number of soft sensors as $16 (4 \times 4)$, $169 (13 \times 13)$, $625 (25 \times 25)$, $1600 (40 \times 40)$, and $2116 (46 \times 46)$. For each of these location sets, the RMSE $\epsilon$ is computed across all estimation locations that are noncoinciding with the soft or hard sensor locations and all time instances for each noise level of the soft data.

Fig. 8 shows the normalized error versus noise standard deviation $\gamma$ for the five soft sensor grid sizes. The normalized error for each $\gamma$ is computed by dividing the RMSE $\epsilon$ by the RMSE obtained for the hard only case, where there are no low-precision sensors deployed, i.e., $N_I = 0$. The normalized error increases with the standard deviation of the noise level of the soft sensors. Furthermore, the normalized error is observed to be lower within the region with a larger number (and, thus, denser array) of soft sensors.

**REFERENCES**


IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING

Timothy C. Havens (S’06–M’10–SM’10) received the Ph.D. degree in electrical and computer engineering from the University of Missouri, Columbia, MO, USA, in 2000.

He is an Assistant Professor in the Departments of Electrical and Computer Engineering and Computer Science and the William and Gloria Jackson Assistant Professor of Computer Systems at Michigan Technological University, Houghton, MI, USA. Prior to his current position, he was a National Science Foundation/Computing Research Association Computing Innovation Fellow at Michigan State University, Lansing, MI, USA. Before working on his Ph.D., he was an Associate Technical Staff at the Massachusetts Institute of Technology, Lincoln Laboratory. His research interests include scalable data mining, network intrusion detection, heterogeneous and big data, social networks, fuzzy sets, sensor networks, and data fusion. He has coauthored over 50 technical publications.

Dr. Havens was the recipient of the Best Overall Paper Award at IEEE International Conference on Fuzzy Systems 2013, the IEEE Franklin V. Taylor Award, and a Best Paper Award from the Midwest Nursing Research Society. He is an Associate Editor of the IEEE TRANSACTIONS ON FUZZY SYSTEMS.

Shanika Karunasekera received the B.Sc degree in electronics and telecommunications engineering from the University of Moratuwa, Moratuwa, Sri Lanka, in 1990 and the Ph.D. degree in electrical engineering from the University of Cambridge, Cambridge, U.K., in 1995. From 1995 to 2002, she was a Software Engineer and a Distinguished Member of the Technical Staff at Lucent Technologies, Bell Labs Innovations, Murray Hill, NJ, USA. In January 2003, she joined the Department of Computing and Information Systems, The University of Melbourne, Parkville, Australia, where she is currently an Associate Professor. Her current research interests include decentralized algorithms for event and anomaly detection, and decentralized data aggregation.

Christopher Leckie received the B.Sc. degree, the B.E. degree (with first class honors) in electrical and computer systems engineering, and the Ph.D. degree in computer science from Monash University, Melbourne, Australia, in 1985, 1987, and 1992, respectively.

He joined Telstra Research Laboratories in 1988, where he conducted research and development into artificial intelligence techniques for various telecommunications applications. In 2000, he joined The University of Melbourne, Parkville, Australia, where he is currently a Professor in the Department of Computing and Information Systems. His research interests include scalable data mining, network intrusion detection, bioinformatics, and wireless sensor networks.


His interests include optimization, pattern recognition, clustering in very large data, co-clustering, and visual clustering. He retired in 2007.

Dr. Bezdek is the past President of the North American Fuzzy Information Processing Society, the International Fuzzy Systems Association (IFSA), and the IEEE Computational Intelligence Society (CIS), the founding Editor of the International Journal of Approximate Reasoning and the IEEE TRANSACTIONS ON FUZZY SYSTEMS, a Life Fellow of the and IFSA, and a recipient of the IEEE 3rd Millennium, IEEE CIS Fuzzy Systems Pioneer, IEEE technical field award Rosenblatt, and the Kampe de Feriet medals.