Stochastic Environmental Research and Risk Assessment (2026) 40:77
https://doi.org/10.1007/s00477-026-03192-4

ORIGINAL PAPER

Real-time contamination monitoring platform using fixed air sensing in underground mining environments

Kate Brown Requist[1] · Moe Momayez[2]

Received: 18 September 2025 / Accepted: 9 February 2026 / Published online: 31 March 2026

Abstract

Airborne contamination has long been acknowledged as a health hazard in underground mining. With a changing regulatory environment in the United States surrounding respirable crystalline silica and other respirable particulates, new tools are required for the assessment of contamination behaviors in underground working environments. Current methods for the simulation and modeling of ventilation systems either lack sufficient spatial resolution or require computational methods that preclude real-time assessment of airborne contamination distributions. Mine ventilation networks are commonly employed for ventilation system design and control, but the one-dimensional graph representation of the network makes contamination behavior within a discrete airway mathematically unobservable. Computational fluid dynamics methods are not suitable for real-time modeling and monitoring despite the superior spatial resolution these methods provide because of the time required for computation and computational barriers to modeling large excavations. We present a real-time contamination monitoring platform leveraging spatial statistics to provide improved spatial resolution compared to mine ventilation network representations while avoiding the computational costs of numerical simulation of the Navier–Stokes equations. Statistical Air Monitoring - Underground (SAM-UG) was developed for the assessment of particulate matter concentration distributions in underground working environments using low-cost, real-time air quality monitors to provide an improved insight into airborne contamination distributions for use in ventilation design and health risk assessment. SAM-UG yielded a total monitoring uncertainty of 30.4% and average RMSE of 0.54μg/m^3 during testing in a portion of an underground mine, similar to existing real-time dust monitoring devices employed in the mining industry.

Keywords Underground mining · Particulate matter · Real-time monitoring · Airborne contamination · Kriging · Manifold embedding · Quasimetric embedding

1 Introduction

Real-time air quality monitoring in underground mines remains an under-explored avenue for improved ventilation engineering and health and safety intervention (Shriwas and Pritchard 2020). In the United States, real-time air quality monitoring is only required in some underground coal mines, which are required to monitor carbon monoxide in certain areas as prescribed under United States 30 Code of Federal Regulations §75.351. These systems are referred to as atmospheric monitoring systems (AMSs), but no similar guidelines exist for other airborne contaminants, nor for other types of underground operations that mine other commodities. Underground metal and nonmetal mines in the United States, at current, have no regulations pertaining to the real-time monitoring of airborne contamination. Nonetheless, with a changing landscape regarding permissible exposure limits for respirable crystalline silica (RCS) and concern about exposure to diesel particulate matter (DPM) in the United States, more considerations should be made to facilitate real-time monitoring of particulate matter concentrations in underground mines.

Real-time monitoring of particulate matter can be achieved using optical particle counting (OPC) technology by estimating mass concentrations based on particle counts across specific size bands. Briefly stated, OPC particulate

✉ Kate Brown Requist
krequist@tnstate.edu

1 Department of Civil and Architectural Engineering, Tennessee State University, Nashville, TN, USA

2 School of Mining Engineering and Mineral Resources, University of Arizona, Tucson, AZ, USA

matter monitors use a laser mounted across from a detector to estimate counts of various particle sizes based on the degree of light scattering measured at the detector (Xu 2015). Various considerations for the use of OPC monitors have been made, especially focusing on their applicability to the mining industry. Adoption of OPC-based monitors are slow; no OPC-based monitors have been approved by the United States Mine Safety and Health Administration (MSHA) for regulatory monitoring. Currently, the only monitor approved by MSHA for regulatory monitoring of particulate matter, specifically respirable coal mine dust (RCMD), is the gravimetric, filter-based Thermo Scientific PDM3700, developed by Thermo Fisher Scientific in Waltham, Massachusetts, USA (Halterman et al. 2018; Volkwein et al. 2006). The Thermo Scientific PDM3700 is notably a personal monitor, which provides a picture of individual exposure to particulate matter, but is sold for upwards of $15,000 USD (Halterman et al. 2018). Non-regulatory area monitors using OPC, such as the Trolex Air XD, are available at a reduced price point, but the $7,000 USD monitor presents an acquisition cost higher than operations may be willing to accept, especially if multiple areas need to be monitored at one time.

Particulate matter encompasses a wide range of human respiratory health hazards of interest in underground mining. Examples of particulate matter can include RCMD, RCS and DPM, among many other species. While these can all be considered under the larger umbrella of particulate matter, current OPC technology is incapable of differentiating across contaminant species and can struggle greatly to accurately count fibrous particles or particles with an aerodynamic diameter smaller than 0.3 μm (Baron 2016). This greatly complicates the adoption of OPC technology for occupational exposure assessment because contaminant species and geometry are the drivers of health outcomes, such as coal worker's pneumoconiosis in the case of RCMD or silicosis in the case of RCS (Donoghue 2004; Laney and Weissman 2014). Nonetheless, OPC technology presents an avenue to provide a first-line indication of airborne particulate contamination as developments in real-time, species-specific particulate matter monitoring technology are ongoing.

Various low-cost real-time OPC monitors are available. These low-cost monitors have similar performance compared to traditional filter and OPC monitors. Ghamari et al. (2022) compared three low-cost OPC particulate matter monitors and found that the accuracy and precision of the low-cost monitors is largely comparable to the accuracy and precision of traditional OPC monitors. Critically, these low-cost monitors have reduced accuracy and precision at higher mass concentrations of particulate matter. Amoah et al. (2023) suggested that low-cost OPC monitors display similar resistances to changes in temperature and humidity as compared to traditional filter and OPC monitors. Wolfe et al. (2024) conducted a one-year field study using low-cost OPC monitors to assess trends in PM_{10} at various locations in the material processing facilities of the mine. This study demonstrated the applicability of low-cost OPC monitoring to mining environments over a long-term installation period with the express focus of exposure assessment and mitigation for mine workers.

The field study from Wolfe et al. (2024) highlights a specific issue with any low-cost area monitoring project. While low-cost real-time OPC monitors are promising, it is still unclear where to place monitors in order to gain the best understanding of contamination dynamics and behaviors. Similarly, there are questions of how many monitors to use (Wolfe et al. 2024). Of the study of underground coal mines in the United States using AMSs conducted by Rowland et al. (2018), the average underground coal mine uses 38 real-time monitors in its AMS. At its core, these open questions indicate a readiness to use low-cost real-time OPC monitors, but there is no general consensus for their application. This leads to concerns about placement and ensuring that observed data accurately reflects contamination dynamics. Requist and Momayez (2025) discusses the placement of air quality sensors in underground mines based on the number of available sensors while remaining cognizant of contamination fate and transport phenomena. In a similar vein, Shriwas and Pritchard (2020) discuss this problem in relation to ventilation monitoring and control, additionally suggesting that this question might be better answered with the development of a mathematical method to estimate contamination concentrations between sensors. A mathematical method to determine concentrations between sensors would provide a clearer picture of contamination dynamics, allowing OPC monitoring to be used to its fullest extent.

Current air quality monitoring paradigms in underground mines can lead to systems that are relatively data-dense but information-poor. Wolfe et al. (2024) yielded large volumes of data, but these data were difficult to interpret as meaningful information for the purpose of contamination distribution monitoring. Environmental monitoring across a variety of contaminant species is commonplace in underground mines, largely as personal monitors or, less frequently, as area monitors. While this can provide large amounts of data, care must be taken to convert those data into meaningful, actionable information (Brown Requist et al. 2023). With a suggestion in Shriwas and Pritchard (2020) to develop a means to estimate contamination concentrations between sensors, fixed air quality monitoring data could provide more information about contamination behaviors. This could instead lead to a real-time monitoring paradigm that returns a distribution of airborne contamination concentrations rather than requiring operations to make ventilation and health and safety decisions based on location-specific point data (Brown Requist et al. 2023).

As a possible answer to the contamination question posed by Shriwas and Pritchard (2020), spatial statistics is a promising method to provide an estimate of contamination concentrations between sensors, and can provide higher spatial resolution than modeling via mine ventilation network (MVN) analysis. MVN models are one-dimensional structures. Because MVNs are represented as graphs, these graphs greatly limit any ability to estimate conditions at different points within an airway. This is because under a graph representation of a mine's ventilation system, airways must be represented as edges of a graph. Using an MVN solver, the edges of the graph are the subjects of simulation. This means that there is no means to achieve a higher spatial resolution than the entire airway. Ultimately, this limits the applicability of MVNs for real-time contamination modeling within discrete airways (Requist et al. 2024; Shriwas and Pritchard 2020). While the graph representation of the ventilation system facilitates fast calculation that is desirable for iterative ventilation design and planning, the lack of spatial resolution makes it unfit for real-time monitoring and modeling of airborne contamination distributions in underground mines. Spatial statistics, however, is not dependent on the graph representation of the mine, and can instead represent the ventilation system as a set of discrete points. This allows for significantly higher spatial resolution, meaning that airborne contamination can be modeled below the airway-level.

Alternatively, computational fluid dynamics (CFD) has the potential to provide much higher spatial resolution, but the computational complexity disqualifies the use of CFD as a real-time monitoring system. Because CFD methods require numerical solutions to the Navier–Stokes equations, the computational overhead of CFD modeling prohibits real-time monitoring. Popular CFD models for indoor particle dispersion modeling include RNG k-ε, Realizable k-ε, and SST k-ω. While CFD models can yield high-quality realizations of air and contaminant behavior in turbulent flow regimes, the corresponding size of these models pose a significant barrier Lawson et al. (2012). Further, readily available and less computationally complex Euler-Euler models struggle with particle dispersion modeling with an inability to consider particle deposition. Euler-Lagrange models perform better with particle dispersion, especially SST k-ω, but are also more computational complex (Zabihi et al. 2025). While demonstrated as somewhat reliable for computational modeling of particulate transport, CFD simulation remains impractical for large-scale simulated environments and impossible for real-time modeling and monitoring.

CFD simulation is notoriously sensitive to grid size (Mora et al. 2002), has limited options for experimental validation (Chang et al. 2020; Duan et al. 2021; Zhang et al. 2021; Zhou et al. 2013), and is incapable of handling simulations at mine-scale with current computational resources (Chang et al. 2020; Duan et al. 2021; Yuan et al. 2016; Zhang et al. 2021; Zhou et al. 2013). Yuan et al. (2016) used CFD for the modeling of mine fires, but struggled to model large mine sections because the size of the model would require an unfeasible amount of time to simulate multiple time-steps in succession. This is similarly discussed in Zhang et al. (2021), where the authors used CFD to model DPM emissions as a means of improving MVN modeling of DPM. The CFD model associated with Zhang et al. (2021) required multiple simplifications to allow for simulation. Naturally, these issues are not conducive to a real-time monitoring system. Similar to CFD, the derivative Fast Fluid Dynamics (FFD) methods are still too computationally intensive to allow for real-time modeling of complex systems like an underground mine's ventilation system. These methods additionally display considerable error when compared to experimental data, raising questions of the reliability and applicability of FFD models (Zheng et al. 2022). Given the space and time complexity of CFD methods, CFD cannot be used for the real-time monitoring of contamination distributions in underground mining environments, and other options must be explored.

Spatial statistics presents an option in between MVN solvers and CFD simulation that provides improved spatial resolution when compared to MVN solvers, and reduced processing time when compared to CFD simulation. The use of spatial statistics stands to provide an estimate of airborne contamination concentrations with higher resolution while remaining fit as a real-time monitoring system. Spatial statistics for the estimation of airborne contamination has been successful in the monitoring of outdoor air quality. Vicedo-Cabrera et al. (2013) used universal Bayesian kriging to estimate environmental exposures to nitrogen dioxide and sulfur dioxide. Before that, Sahu and Mardia (2005) used a Bayesian kriged Kalman model to forecast airborne contamination in an urban area. Somewhat more recently, Beauchamp et al. (2017) and Beauchamp et al. (2018) explore different methods to improve the performance of statistical air quality models. The methods presented in Beauchamp et al. (2017) and Beauchamp et al. (2018) make use of various detrending methods. Beauchamp et al. (2017) modifies the distance metric used in kriging. Beauchamp et al. (2018) detrends the input kriging data by estimating traffic contributions to air quality.

Brown Requist and Momayez (2024a) focused on a small-scale test with various ordinary kriging methods and spatial structures using synthetic data for the assessment of carbon monoxide for underground coal mining. This work showed that ordinary kriging is a promising avenue for real-time contamination distribution modeling in underground mining environments, but more work is needed to validate the performance of spatial statistics with experimental data. Brown Requist and Momayez (2024a) showed that spatial statistics should theoretically be capable of assessing contamination distributions, but the overall workflow required an extensive understanding of spatial statistics to construct estimations, such as the selection of model variogram equations. Additionally, there is a need to develop a definitive estimation workflow and software platform to facilitate real-time monitoring based on spatial statistical estimation of airborne contamination distributions with fixed air quality monitoring. This definitive workflow and software platform should ideally remove the need for an in-depth understanding of spatial statistics by way of automated determination of covariance structures, model validation, and contamination distribution estimation.

We present an initial attempt at a real-time monitoring platform that utilizes multiple variogram equations and spatial realizations of mine geometry. This platform, referred to as Statistical Air Monitoring - Underground (SAM-UG) was developed for use in a portion of an underground mining research facility in Sahuarita, Arizona. SAM-UG was originally designed to monitor particulate matter but the statistical basis for estimation makes the platform transferrable to other airborne contaminants. This platform uses low-cost OPC monitors to estimate distributions of $PM_{2.5}$ across a portion of an underground mine excavation using fixed sensing data obtained from monitors installed on the ribs of the excavation. This initial platform will serve as a basis for additional statistical and computational geometric developments to improve airborne contamination distribution estimation quality in SAM-UG.

2 Methods

SAM-UG leverages spatial statistics for the estimation of real-time airborne contamination distributions using fixed air quality sensing. Developed in C++, estimated distributions are constructed using ordinary kriging across a variety of spatial realizations and three model variogram equations.

2.1 Preprocessing of spatial structure

To facilitate kriging, distances between fixed air quality monitors and estimation locations within the underground working environment must be calculated to determine the relevant kriging weights. These distances can be calculated in preprocessing because these distances do not meaningfully change unless the excavation itself changes. These distances are calculated under two distinct spatial structures.

Linear (Euclidean) distances are straightforward to calculate as per Eq. 1:

$$d = ||b - a|| \tag{1}$$

where d is the Euclidean distance between points a and b (in local mine coordinates) as calculated via the L2 Norm.

Naturally, linear distances do not have any means to account for an excavation's geometry, meaning that distances measured may assume that contamination behavior is unimpeded by the presence of rock or other solid interfaces such as stoppings or seals. Some other metric could be constructed to account for excavation geometry and airflow behaviors to better reflect spatial variability. Brown Requist and Momayez (2024b) details the construction of a surrogate distance value that respects air flow direction and excavation geometry. This method is used to account for airborne contamination behaviors by assigning distance values based on the direction of contaminant transport as mediated by the mine's ventilation system. Formally stated, the surrogate distance value calculation method presented in Brown Requist and Momayez (2024b) describes a quasimetric embedding (QE) of the mine ventilation system. This embedding preserves effects of air flow and excavation geometry which can be used to approximate a manifold over which kriging can proceed. The resulting space retains a euclidean distance, but additionally has a quasimetric that accounts for the effective distance between locations within the excavation when considering airflow and geometry. This effective distance value is not a proper metric, as it does not display symmetry. In essence, the distance between a point A and another point B is not identical to the distance between B and A at all points within the manifold (Wilson 1931). For example, if A is upwind of B, the effective distance from A to B should be smaller than the effective distance from B to A because airborne contamination will preferentially flow along the direction of airflow. Practically, this is most similar to the use of non-linear distances for kriging or shortest-path distances for ordinary kriging with locally varying anisotropy (Boisvert and Deutsch 2008; Ver Hoef 2018).

While linear distances lead to kriged estimations which consider a straight line relationship governing data covariance, this does not yield a realistic representation of data conditions, as airborne contaminants cannot travel through rock. The surrogate distance value calculated to account for airflow is used as the QE distance which then facilitates kriging using the QE framing. The use of QE distances allows for the assessment of data covariance (and subsequent modeling of the covariance function) such that airflow is respected. In practice, moving from an upwind position to a downwind position is, in essence, a "shorter"

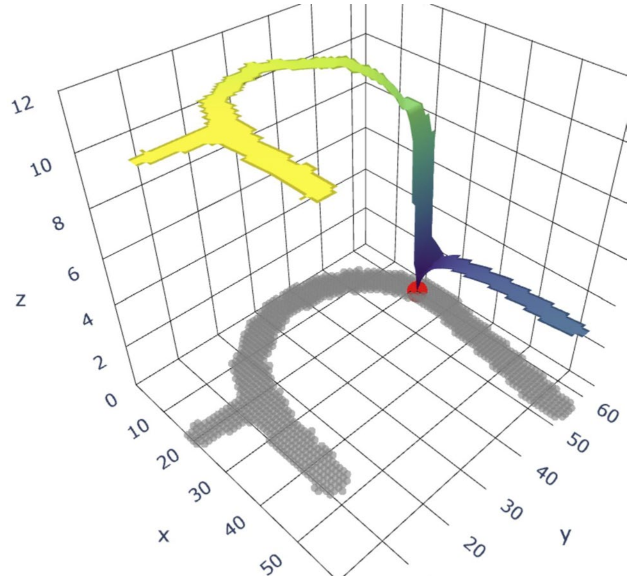

Fig. 1 Visualization of the latent space presented by the quasimetric embedding. The vertical axis represents the relative distance as measured from the red dot to all other parts of the mine (Euclidean projection in gray). All coordinate values are relative to preserve aspect ratio and not indicative of any specific distances or units

distance than vice versa. This means that upwind-to-downwind movements are measured as significantly smaller distances than downwind-to-upwind movements. Figure 1 highlights this relationship, with the vertical axis indicating the QE distance as measured from one point (red dot) to all other points within the mine. This is an attempt to visualize the QE of the ventilation system, but is a small view of the latent space, and this representation of the latent space holds for only the single point. Each point within the ventilation system has its own set of QE distances from that point to every other point in the system. As such, it is impractical to attempt to visualize the entirety of the QE, as this would require at minimum $2n^2$ dimensions to faithfully represent in a Euclidean framing.

Linear and QE distances were calculated using Flowstar version 0.2.1, a free computer program developed following methods outlined in Brown Requist and Momayez (2024b). Flowstar facilitates the modeling of airflow and construction of the QE so that QE distances can be used for kriging. Preprocessing of these distances removes the need for SAM-UG to calculate these distances as a part of the estimation workflow, reducing the time needed to produce a viable contamination distribution estimate. This preprocessing step is conducted using Flowstar to arrive at the QE distances which are then used to construct the required covariance structures. These QE distances are not used as kriging weights directly, but are instead used to define the covariance function used in the kriging system of equations that ultimately result in the kriging weights for estimation.

Fig. 2 PurpleAir Classic air quality monitor installed in the underground mining facility connected to a rechargeable 20 Ah battery

By considering linear and QE distances within the SAM-UG platform, the best spatial structure can be selected to yield an estimated contamination distribution with the smallest estimation uncertainty.

2.2 Placement of air quality sensors

Fifteen PurpleAir Classic air quality monitors (PurpleAirs) developed by PurpleAir, Inc in Draper, Utah, USA, were placed in a portion of an underground mining research facility in Sahuarita, Arizona. The PurpleAir Classic uses two PMS-5003 laser particle counters manufactured by Plantower Technology in Nanchang, Jiangxi, China with a counting efficiency of 98% and a maximum consistency error of 10%, yielding a total sensing uncertainty of approximately ±12%. Each PurpleAir was mounted to an ABS enclosure and powered with a rechargeable 20 Ah battery (Fig. 2). This allows for a continuous monitoring time of approximately 100 h before the battery must be recharged. Monitor installation locations were determined using EPAQS version 0.3.1, based on the Efficient Placement of Air Quality Sensors (EPAQS) algorithm as presented in Requist and Momayez (2025). This algorithm considers the QE distance between locations within the excavation to return a set of sensor spacings that minimize the clustering of input data. This is vital to promote the numerical stability of the kriging equations and eliminate the need to decluster already sparse fixed monitoring data. In traditional geostatistics, data declustering is a standard part of the estimation workflow, generally involving the use of declustering weights to minimize under- or over-representation of the estimation domain and to minimize the risk of singular covariance matrices (Bourgault 1997; Deutsch 1989). Without proper data spacing with regards to QE distances, QE-based kriging models have a higher chance of returning negative kriging variances due to under- or over-representation of parts of the domain. Further, this would dramatically increase the risk of singular covariance matrices when estimating with QE distances. By spacing sensors according to QE distances, this works to improve the numerical stability of the kriging equations by minimizing data clustering.

Figure 3 displays the sensor installations as recommended by EPAQS version 0.3.1, with installation locations constrained to the ribs of the excavation to avoid installation requiring hanging monitors from the back of the excavation. Sensors were installed at 62(±6) inches (approximately 157 cm) above the floor of the excavation to approximate the breathing zone of mine workers following the anthropometrics study conducted by Gordon et al. (2014). Sensor placement as guided by EPAQS was constrained to the ribs (walls) of the excavation to allow for operation of equipment. This leads to asymmetry of the sensor layout, with sensors 7, 11, and 15 placed on the opposite rib than the others. From an operational feasibility standpoint, installation of sensors hanging from the back (ceiling) of the excavation is not possible, although the spatial distribution of sensors could perhaps be slightly improved.

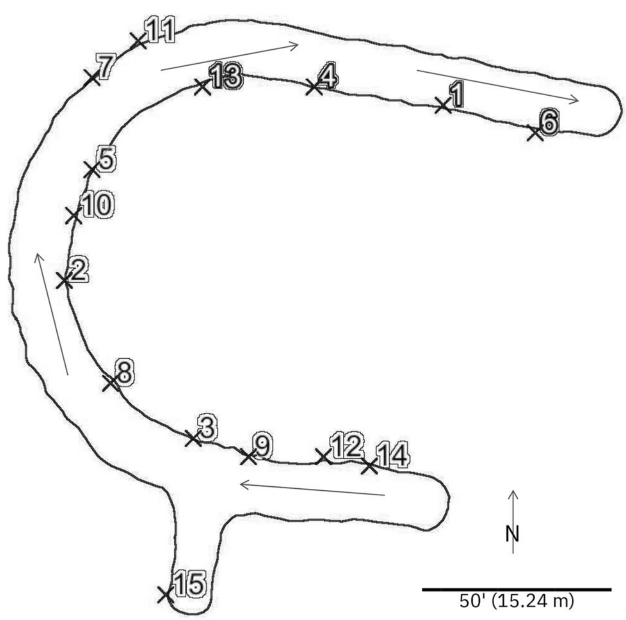

Fig. 3 Placement of 15 $PM_{2.5}$ sensors using EPAQS version 0.3.1. Sensor locations are numbered in accordance with unique sensor IDs assigned by the EPAQS software for integration with SAM-UG. The sensor installed in location number 1 as indicated in the graphic was subsequently removed from service due to anomalous sensor values

2.3 Data acquisition and variogram model selection

Real-time data collection is feasible in one-minute intervals via the PurpleAir API. The resulting $PM_{2.5}$ data is passed into SAM-UG as a .json file. SAM-UG will accept any .json file of air quality data, making it convertible between various real-time monitoring APIs or data streams and contamination species. Once data has been acquired, SAM-UG will automatically select the best performing variogram model and proceed with estimation via ordinary kriging.

Model variogram selection considers three spatial paradigms and three variogram equations. Spatial paradigms used in kriging include linear distances, QE distances, and QE distances that have been conditioned with linear autocorrelation information. Linear distance is arguably the most widely-used spatial paradigm for ordinary kriging seen in modern geological resource modeling, but naturally cannot account for an excavation's geometry or airflow. QE distances respect airflow and geometry, but can struggle with numerical stability (Brown Requist and Momayez 2024a). As an additional method, QE distances can be conditioned using autocorrelation observed in the Euclidean space. This method was originally suggested by Ver Hoef (2018) for kriging over linear networks. Following Ver Hoef's application to non-euclidean distances, the method can be rationally extended to use QE distances. This method yields a covariance matrix that is typically positive-definite (PD), constituting a permissible variogram (Armstrong and Jabin 1981; Goovaerts 1997; Webster and Oliver 2007) that is not feasible using QE distances because the resulting covariance matrix is necessarily non-symmetric. The covariance structure modified from Ver Hoef (2018) is displayed in Eq. 2,

$$\mathbf{\Sigma} = \sigma_p^2 \mathrm{R}_{\mathrm{A},\alpha} \mathrm{R}_{\mathrm{D},\eta}^{-1} \mathrm{R}_{\mathrm{A},\alpha}^{T} + \sigma_0^2 \mathrm{I} \tag{2}$$

where $\mathbf{\Sigma}$ is the conditioned covariance matrix, R is an autocorrelation matrix based on a QE distance matrix A or linear distance matrix D with QE range parameter α or linear range parameter η. The partial sill, σ_p^2, and nugget σ_0^2 can be determined via restricted maximum likelihood (REML) alongside range parameters α and η. Alongside linear and quasimetric (QE) spatial paradigms, the Ver Hoef (2018) method as modified for use in SAM-UG is henceforth referred to as the "conditioned" spatial paradigm for the sake of simplicity.

This conditioned paradigm uses linear autocorrelation, $\mathrm{R}_{\mathrm{D},\eta}$, as the core of the covariance structure, and the QE autocorrelation, $\mathrm{R}_{\mathrm{A},\alpha}$, conditions the underlying linear autocorrelation with quasimetric spatial information. This helps to strike a balance between a purely linear or purely quasimetric estimation domain. This is especially desirable when the linear domain and quasimetric domain do not share the same scale. By way of the unified covariance structure presented in Eq. 2, range parameters α and η serve to address these differing scales when combining the two spatial structures as part of the conditioned paradigm. Essentially, this approach can be thought of as adding an additional design matrix (quasimetric autocorrelation) to the initial design matrix (linear autocorrelation) to arrive at a covariance matrix which blends both spatial paradigms.

The three variogram equations employed by SAM-UG include the spherical, exponential, and Matérn semivariance functions (Eqs. 3, 4, and 5, respectively). These variogram equations were selected for multiple reasons. The spherical variogram equation is only valid for estimation domains over $\mathbb{R}^n$ with $n \leq 3$ (Armstrong and Diamond 1984; Pyrcz and Deutsch 2006). While this remains appropriate for kriging over a linear estimation domain, this is not guaranteed for kriging over the QE domain. This is because the QE is an embedding of the euclidean manifold (in this case as some subset of $\mathbb{R}^2$) into the quasimetric manifold, which constitutes a Hilbert space, essentially subset of $\mathbb{R}^n$ such that $0 < n < \infty$ (Son 1982). Unless the Hilbert space can be reanalyzed as a domain over $\mathbb{R}^m$ with $m \leq 3$, the spherical variogram is not technically valid. In this case an *n*-spherical variogram equation should be used, but the value of *n* is not mathematically observable, and may not be finite. The exponential variogram was selected because it remains explicitly defined for all dimensions (Oliver 1995), addressing the issue of uncertainty of the resulting Hilbert space. The Matérn variogram was likewise chosen for this reason, with the added benefit of the Matérn variogram's competence to capture variability at differing scales (Marchant and Lark 2007). By considering these three variogram equations, SAM-UG has more freedom to select an appropriate variogram when observations are highly variable between time steps.

All variograms were fit via REML and Sequential Least SQuares Programming (SLSQP) using Moore-Penrose matrix pseudoinversion to automate the variogram fitting process and to improve numerical stability of covariance matrices that are not PD and potentially singular during inversion. Strictly speaking, the Moore-Penrose pseudoinversion of a singular covariance matrix conditions the (pseudo-) inverted matrix to return a least-squares approximation of the kriging system of equations. While this does have some implications for the reliability of kriging performance metrics (i.e., kriging variance) that warrant discussion, the overall merits of the use of Moore-Penrose pseudoinversion for kriging over singular covariance matrices lies outside the scope of this work. REML is a favored method for the estimation of covariance structures because it is relatively robust to outliers, removes the need to arbitrarily bin distances for the construction of variograms, and can fit model covariance functions to experimental data without first plotting an experimental semivariogram (Cressie 1991; Cressie and Lahiri 1996; Heyde 1994; Marchant and Lark 2007; Ver Hoef 2018). In effect, REML can be used to automate the fitting of covariance structures based on experimental data. Via leave-one-out cross validation (LOOCV), the best model variogram is selected based on root mean square error (RMSE), proportion of time that the 90% confidence interval covers the true value, and the number of negative kriging variances observed during validation.

$$\gamma(h) = \begin{cases} \sigma_0^2 + \sigma_p^2 \left(\frac{3h}{2a} + \frac{h^3}{2a^3} \right) & \text{if } 0 \leq h \leq a \\ \sigma_0^2 + \sigma_p^2 & \text{if } h > a \end{cases} \tag{3}$$

$$\gamma(h) = \sigma_0^2 + \sigma_p^2 \left(1 - e^{\frac{-h}{a}} \right) \tag{4}$$

$$\gamma(h) = \sigma_0^2 + \sigma_p^2 \left(1 - \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{h}{a} \right)^{\nu} K_\nu \frac{h}{a} \right) \tag{5}$$

where $\gamma(h)$ is the semivariance at lag distance *h*, *a* is the range parameter, ν is the Matérn smoothness parameter, and K_ν is the modified Bessel function of the second kind of order ν. The variogram parameters (i.e., σ_0, σ_p, ν, and range parameter *a*) are determined by optimization specific to observed data, and these parameters hold for the entirety of the estimation domain. This represents the entirety of parameters that must be determined, except for the case of conditioned models, which have a QE range parameter as well as a Euclidean range parameter.

The best model variogram should ideally have low RMSE, a high coverage of the 90% confidence interval, and few negative kriging variances. These performance statistics are ranked to select the best model variogram (Eq. 6).

$$\gamma_{opt}(h) = \operatorname{argmin}(\operatorname{rank}(RMSE) + \operatorname{rank}(C_{90\%}) + \operatorname{rank}(NKV)) \tag{6}$$

where $\gamma_{opt}(h)$ is the preferred variogram equation (covariance structure) as determined by the proportion of times the 90% confidence interval includes the true value ($C_{90\%}$), the number of negative kriging variances obtained during LOOCV (*NKV*), and RMSE. A new variogram equation must be constructed for each time step for each possible model combination, because observed data covariance is required for successful kriging. As sensor data changes, data

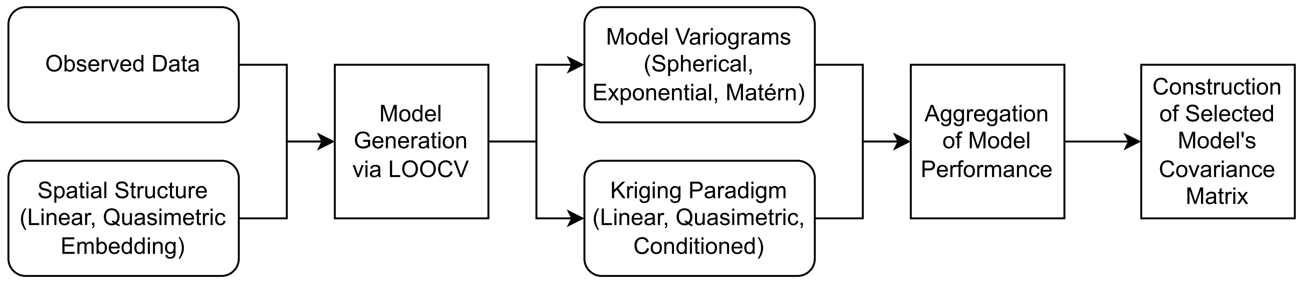

Fig. 4 SAM-UG model selection workflow using REMLengine version 1.7

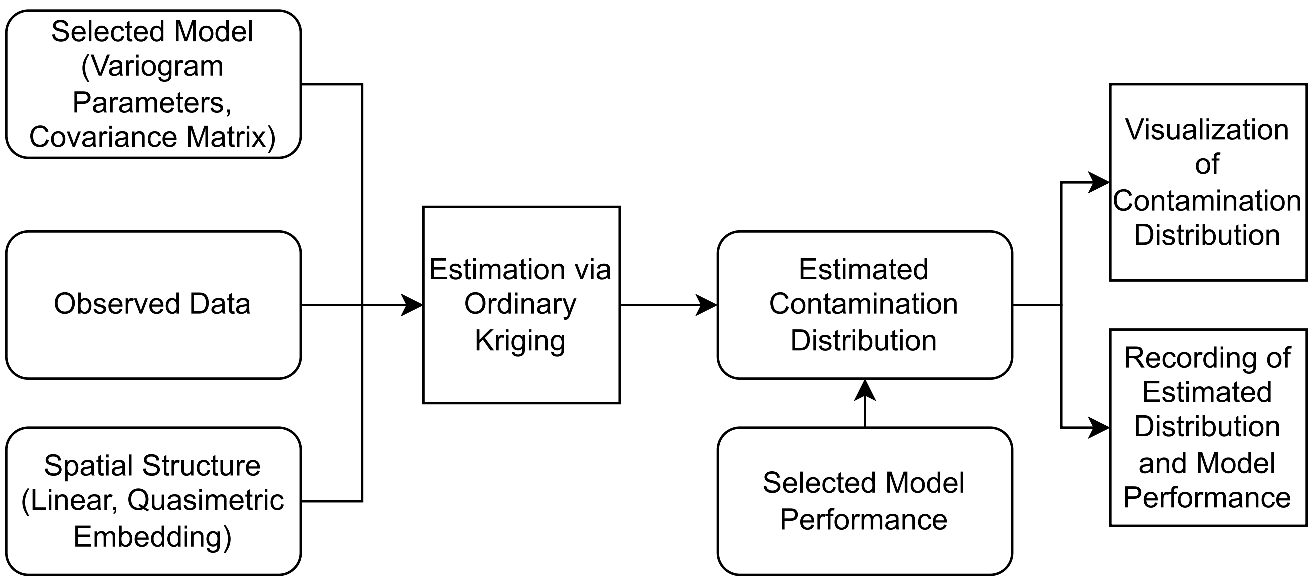

Fig. 5 Estimation and visualization workflow in SAM-UG

covariance likewise changes, meaning the variogram must be updated. All models were fit, validated, and selected using REMLengine version 1.7, a C++ library developed by the corresponding author to facilitate covariance structure fitting across various non-traditional, user-defined spatial structures and variogram equations.

Figure 4 displays the model selection workflow, starting with the observed real-time data and spatial structures as generated by Flowstar version 0.2.1. Models are generated under LOOCV across the three spatial paradigms and three variogram equations, yielding a total of $9N$ covariance matrices for N monitors to assess model performance. LOOCV model performance (RMSE, $C_{90\%}$, and NKV) is aggregated, returning a unified covariance matrix and reporting the relevant variogram parameters, variogram equation, spatial paradigm, and model performance statistics. This aggregated performance data is then used as per Equation 6 for model selection. By automating the model selection process via REMLengine 1.7, SAM-UG only displays an estimate made with the best-performing covariance structure. Further, automated fitting of the covariance structure and model selection removes the need for extensive spatial statistics and mine ventilation knowledge, making SAM-UG approachable for users like industrial hygiene or mine health and safety professionals. Model performance statistics are continuously logged in SAM-UG for post-hoc assessment of model performance. These statistics are also presented to the user during visualization of the contamination distribution after estimation via ordinary kriging.

2.4 Contamination distribution estimation and visualization

Contamination distributions are estimated via ordinary kriging, which uses the covariance matrix and corresponding variogram equation to generate the best linear unbiased estimator (BLUE) to estimate contamination concentration at a discrete location as a linear combination of observed $PM_{2.5}$ mass concentrations from fixed sensors. The resulting distribution is displayed as a two-dimensional plot that reports model performance and updates in one-minute intervals or as new data is acquired. Figure 5 outlines this workflow from the selected model to the visualization of the contamination distribution.

2.5 Testing of SAM-UG

SAM-UG was deployed over a four-hour period in a portion of an underground mining research facility in Sahuarita, Arizona with 15 PurpleAirs with the intention of monitoring $PM_{2.5}$. During the testing period, light vehicle and pedestrian activity was intermittent, providing variability in contamination generation conditions across the study period. Of the 15 installed PurpleAirs, Sensor 1 (with location indicated in Fig. 3) was removed from consideration in the study dataset due to anomalous sensor readings, typically an order of magnitude higher than all other sensor readings. Without a clear indication of secondary particulate entrainment at the sensor location, data were assumed erroneous and therefor omitted. The resulting spatial gap in data was deemed insignificant, as sensor placement with EPAQS is designed to reduce data clustering. Ideally, a properly functioning sensor would be beneficial, but the uncertainty of estimates for sensor 6 (the downwind sensor) was comparable to the uncertainty of estimates for the other sensors during LOOCV. The omission of this sensor had no observable impact on the linear independence of covariance matrices, in part due to fitting variogram equations using REML. Future safeguards should be developed within the software to identify potentially faulty sensors and notify the user to decide whether to omit the data from estimation. Figure 6 displays $PM_{2.5}$ values measured by the remaining 14 OPC monitors over the four-hour period with a total of 247 observations in one-minute intervals. These sensor data display the average $PM_{2.5}$ concentration over the one-minute interval in $\mu g/m^3$. These data are not directly displayed in the SAM-UG software in the current version, but future updates will include a real-time graph of all sensor data. This will further facilitate the identification of erroneous data. Model performance was continually logged by SAM-UG, and the resulting performance was aggregated to assess the long-term performance of the monitoring platform.

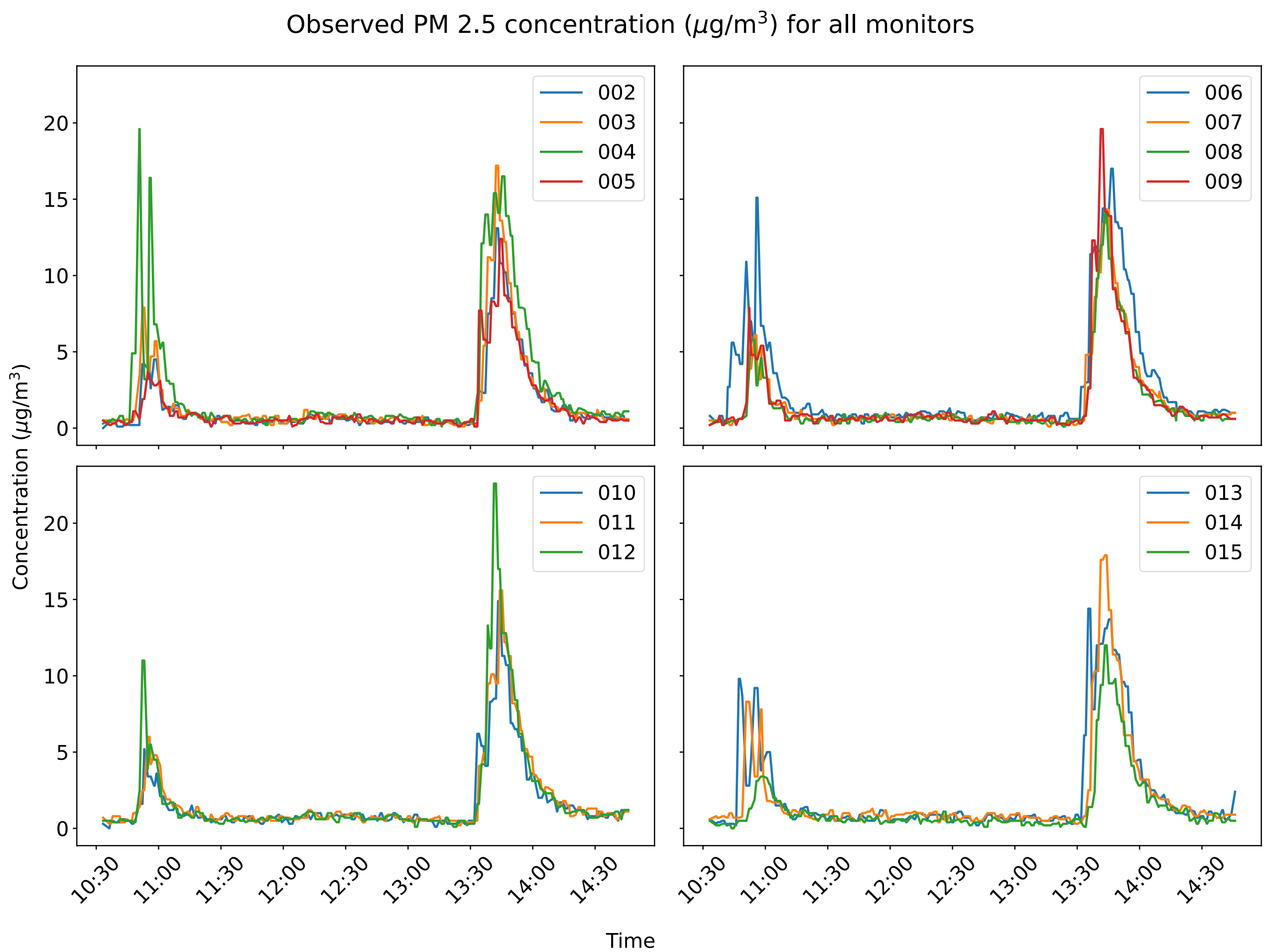

Fig. 6 $PM_{2.5}$ mass concentrations ($\mu g/m^3$) for all PurpleAirs over the four-hour study period. Data have been divided into subplots for the sake of readability.

3 Results and discussion

SAM-UG uses a command-line interface (CLI) (available in Supplementary Information), and requires four configuration files alongside a data folder or other server connection. The four file types required by SAM-UG are .LDM (linear distance matrix), .QEM (quasimetric embedding matrix), .LMP (linear map), and .QMP (quasimetric map). These four files are automatically generated by EPAQS during sensor placement and are then passed to SAM-UG via the CLI. Users can indicate the type of data connection (static file folder or dynamically updating folder of .json data) and plotting utility (real-time in a separate window or saved as a .pdf or .png file). Estimated contamination distributions are saved as .DIST files to archive estimates, regardless if plots are explicitly saved, and performance statistics are saved as a single .STATS file when the application is closed.

SAM-UG was trialed over a four-hour period (a total of 247 observations) using data from 14 PurpleAirs to measure $\mathrm{PM}_{2.5}$ mass concentration in one-minute intervals. Linear distances were preferred for estimation using SAM-UG, which selected a model variogram based on linear distances 78.5% of the time (n=194), followed by selection rates of 12.6% and 8.9% for quasimetric and conditioned models, respectively. Table 1 highlights the selection frequency and model performance for all nine model types. Negative kriging variances were not observed for any models across the study period, which has been raised as a potential issue while kriging using non-linear distances, albeit with varied findings (Brown Requist and Momayez 2024a, 2025; Ver Hoef 2018; Boisvert and Deutsch 2008; Davis and Curriero 2019; Gardner et al. 2003; Little et al. 1997; Rathbun 1998). This finding closely matches previous assessments of the estimation of contamination distributions in mine air, and we suggest that this outcome may be a result of sensor placement guided by quasimetric distances, the use of SLSQP for fitting model variograms with REML, and, of course, the model selection criteria themselves, which penalize the presence of negative kriging variances during LOOCV.

Table 1 Performance of all SAM-UG model types

Equation	Paradigm	*n* (% Frequency)	Uncertainty (%)	RMSE (μg/m^3)	$C_{90\%}$
Exponential	Linear	33 (13.36)	24.6	1.19	0.88
Exponential	Quasimetric	6 (2.43)	26.8	0.84	0.80
Exponential	Conditioned	0 (0)	–	–	–
Spherical	Linear	27 (10.93)	27.6	0.79	0.85
Spherical	Quasimetric	4 (1.62)	25.7	0.67	0.89
Spherical	Conditioned	0 (0)	–	–	–
Matérn	Linear	134 (54.25)	32.2	0.37	0.84
Matérn	Quasimetric	21 (8.50)	31.2	0.44	0.86
Matérn	Conditioned	22 (8.91)	32.6	0.33	0.87

Assessment of the kriged models using LOOCV shows a reasonable internal consistence of the resulting estimates, but more work is needed to establish external physical validity. Future work in this area will require extended studies. One potential avenue to assess physical validity would be comparison with CFD simulation for a variety of mine ventilation systems and layouts. CFD methods are reliable, and benchmarking model validity against methods that are known to be reliable will help to verify the performance of SAM-UG for widespread use in mining operations. Work for further validation is ongoing alongside the development of additional physics-informed statistical approaches.

Linear models are selected more frequently in SAM-UG when compared to the frequency of quasimetric and conditioned models. This paradigm is simpler than either the quasimetric or conditioned paradigms, which may play a role in explaining the discrepancy in selection rate. The relatively simple layout of the study area may contribute to this outcome, which may lead to the higher selection rate of linear models compared to quasimetric and conditioned models. Considerations should be made about the potential drawbacks of the use of linear models, especially because they inherently assume that data covariance behaves linearly without respect to mine geometry or airflow. Further, the frequency of Matérn variogram equations is unsurprising, as the smoothness parameter ν can help to address spatial processes with differences in local (small scale) behavior compared to larger-scale covariance (Marchant and Lark 2007). Outcomes for quasimetric models, especially using Spherical and Matérn variogram equations further highlight that quasimetric models can outperform linear models, even outside of the model selection criteria presented in Eq. 6. When considering the sensor data presented in Fig. 6, quasimetric and conditioned spatial paradigms are selected with higher frequency during instances of high data variability (i.e., 10:30–11:15 and 13:30–14:15). This suggests that these quasimetric and conditioned paradigms may be more suitable for highly variable or even non-stationary data, which follows closely with other findings related to applications of non-linear distances for kriging (Fouedjio 2017; Varouchakis et al. 2025; Martin et al. 2019). Considerations should be made for the inclusion of the conditioned paradigm in future iterations of the SAM-UG software because of the low selection rate. This would reduce the number of variogram models that must be assessed, but we feel that the overall computational benefit is low, considering the linear relationship of time complexity with this part of the model selection process. More investigation in this area is warranted, especially because mining environments are highly dynamic, and second-order stationarity cannot be guaranteed. Within the scope of this paper, all data were assumed to display second-order stationarity for the purposes of geostatistical modeling, and the lack of negative kriging variances for selected models suggests the presence or absence of second-order stationarity is not a major hindrance in the application of the SAM-UG platform.

There exists a higher likelihood negative kriging variances during estimation when employing non-Euclidean distances or quasimetric models (Armstrong 1992; Davis and Curriero 2019; Ver Hoef 2018). Because the corresponding covariance matrices are inherently non-symmetric in quasimetric models, they do not constitute permissible variograms (Armstrong and Jabin 1981; Armstrong 1992; Goovaerts 1997), *per se*. With the use of the Moore-Penrose pseudoinverse, singular quasimetric covariance matrices are effectively invertible. It should be noted that in the development and testing of SAM-UG, it appears that quasimetric covariance matrices are strictly non-singular, likely due to the use of a sensor spacing paradigm based on quasimetric distances rather than Euclidean distances (Brown Requist and Momayez 2025). Even if the covariance matrix is non-singular (invertible) or PD, negative kriging variances can still arise during estimation. By using the Moore-Penrose pseudoinverse, all non-invertible matrices yield a least-squares approximation of the linear kriging system. Negative kriging variances can occur for myriad reasons, including the spatial configuration of data and the variogram structure (Armstrong 1992). Although the spatial configuration of data remains constant, variogram structures are continually changing across the four-hour study period because of changes in input (observed concentration) values. This is additionally complicated by the lack of PD or positive-semidefinite (PSD) covariance matrices when using quasimetric distances. It is somewhat surprising to note the absence of negative kriging variances in the face of the numerical stability difficulties posed by kriging over quasimetric estimation domains. The model selection criteria (Eq. 6) certainly play a large role in the presence of negative kriging variances in selected models, especially because of the penalty assigned to negative kriging variances during selection via LOOCV.

Critically, REML optimization approach plays a role in the numerical stability (and therefore reliability) of model covariance and the kriging system of equations. SLSQP was selected as the non-linear optimization algorithm for the fitting of model covariance equations using REML, which we have found to be the superior REML optimization algorithm in the course of developing the REMLengine library and SAM-UG application. Negative kriging variances stemming from quasimetric or conditioned models were encountered frequently using alternative non-linear optimization algorithms such as Constrained Optimization BY Linear Approximation (COBYLA), Bound Optimization BY Quadratic Approximation (BOBYQA), and Nelder-Mead Simplex. While COBYLA, BOBYQA, and Nelder-Mead Simplex are derivative-free optimization methods, SLSQP requires construction of a gradient by way of finite differences (Conn et al. 1997; Kraft and Schnepper 1989). As a result, SLSQP requires more matrix inversions within the objective functions than derivative-free methods. In practice, this means that SLSQP is theoretically more computationally complex than derivative-free methods, but we have observed that SLSQP tends to arrive at a more stable model covariance function (no negative kriging variances) faster than COBYLA, BOBYQA, and Nelder-Mead Simplex.

This behavior is likely due to the nature of the loss surface with respect to model covariance variables. In the case of REML, this loss surface is the negative log-likelihood as a function over the model covariance parameter space (e.g., nugget effect, total sill, range). When comparing the various spatial paradigms and various model covariance equations, the loss surface is likely to be very noisy and highly variable when considering a spatial structure that remains constant while observed data are changing over the study period. As an optimization method that requires a gradient, SLSQP has access to the nature of the loss surface, which may explain why models fit with SLSQP are less likely to return negative kriging variances when compared to derivative-free methods. With extremely noisy loss surfaces, derivative-free methods like COBYLA, BOBYQA, and Nelder-Mead Simplex are more likely to terminate in some potentially unstable local minimum of the loss surface (Kraft and Schnepper 1989). More work is needed to qualify the nature of these loss surfaces, including in Euclidean space. With an improved understanding of loss surface behavior during REML fitting of covariance functions, the overall quality of model fit stands to greatly improve.

The lack of PD or PSD covariance matrices when attempting to krige contamination distributions using a quasimetric spatial paradigm is a problem that should eventually be addressed. Davis and Curriero (2019) suggest that efforts should be taken to estimate the covariance structure over the non-Euclidean domain before pursuing post-hoc corrections to covariance matrices such that they are PD or PSD. The ultimate solution to this issue lies outside of the scope of this paper, but approximation by spectral decomposition could be used to arrive at a PSD covariance matrix. Because the QE distances used for estimation are not symmetric, care must be taken to develop an appropriate method to symmetrize the deficient covariance matrix in some manner that still maintains some semblance of the underlying spatial structure governed by the mine's ventilation regime. Importantly, this is not a consideration that has been widely made when considering the application of non-Euclidean distances in geostatistics, with prior work on non-Euclidean distances involving metric embeddings (symmetric covariance matrices) rather than quasimetric embeddings (asymmetric matrices). While it would be feasible to construct a PSD covariance matrix based on QE distances with approximation by spectral decomposition, this additional operation would add a layer of computational complexity that may not be suitable for a real-time system. Considering an absence negative kriging variances during estimation with QE distances in SAM-UG, we believe that this issue does not represent a massive barrier to the implementation of the SAM-UG platform, especially in light of the application of the model selection criteria which penalize the presence of negative kriging variances.

Because SAM-UG dynamically selects models based on performance statistics, measures like RMSE cannot be used to compare inter-model performance. As such, Table 1 includes the average estimation uncertainty of each model type as a percentage of the range of observed values for each respective model. The overall spread of these estimation uncertainties (which include the ±12% PurpleAir sensor uncertainty) is low. Across the four-hour study period, SAM-UG had an average estimation uncertainty of 27.9% when excluding PurpleAir sensor uncertainty and 30.4% including sensor uncertainty. SAM-UG showed an average RMSE of 0.54 μg/m^3, and an average $C_{90\%}$ of 0.85. Although Matérn-Linear models were most frequently chosen (n=134), the dynamic model selection within SAM-UG leads to meaningful improvements in overall performance, displaying reductions in estimation uncertainty and RMSE.

Estimated contamination distributions are visualized using the low-level plotting utility GNUplot (Figs. 7, 8, and 9), displaying the timestamp of the input data, spatial paradigm, variogram equation, RMSE, and coverage of the 90% confidence interval. The GNUplot plotting utility is used to display the mine footprint oriented to the input coordinate system used for mine survey, with the color bar indicating the estimated concentration of $PM_{2.5}$.. These plots can be shown in real-time (updating in one-minute intervals corresponding with new incoming sensor data) within SAM-UG or saved in .png or .pdf format based on user selection via the CLI when starting the SAM-UG program. The plotting function in SAM-UG currently lacks the sophistication for broad applicability to other contamination types, a limitation that is expected to be resolved with future platform architecture developments, specifically the development

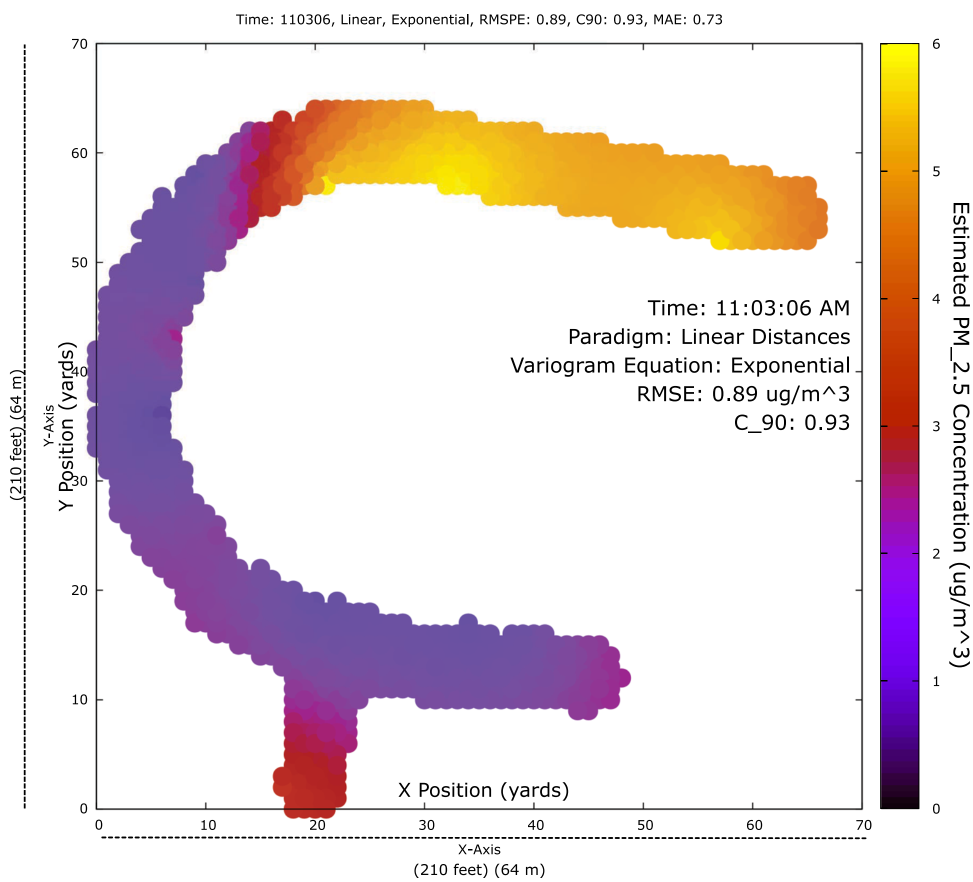

Fig. 7 Estimated contamination distribution at 11:03 AM using Exponential-Linear model. The plot as generated by SAM-UG has been altered in post-processing to provide clearer and more descriptive labels.

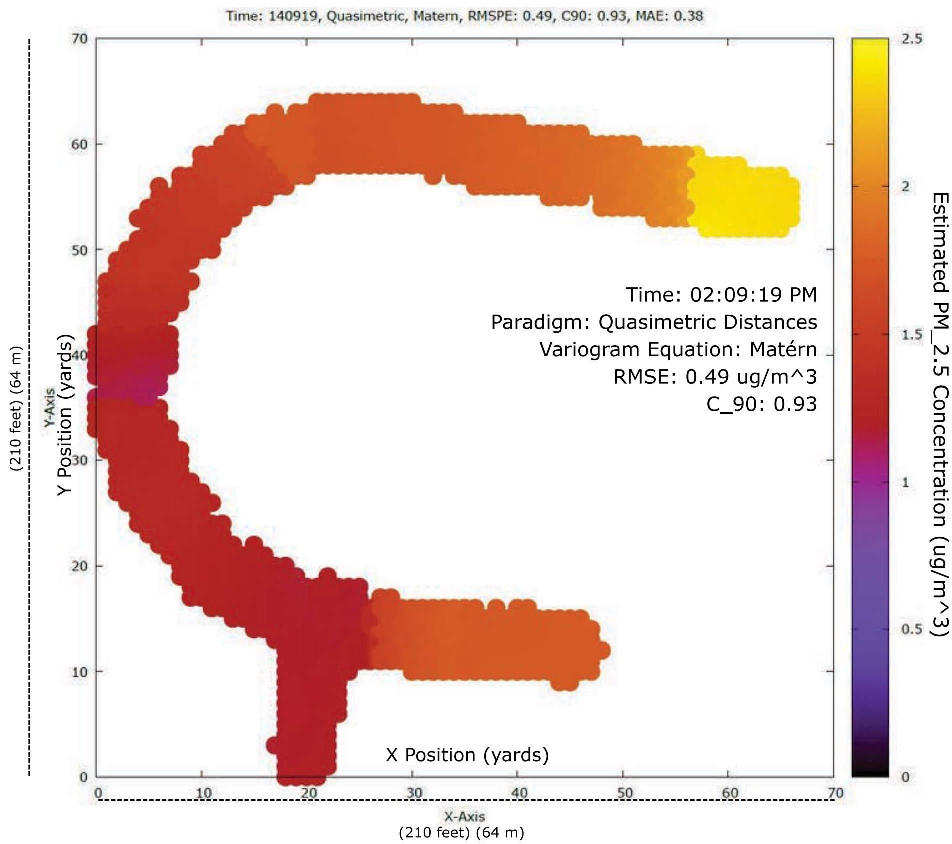

Fig. 8 Estimated contamination distribution at 2:09 PM using Matérn-Quasimetric model. The plot as generated by SAM-UG has been altered in post-processing to provide clearer and more descriptive labels.

of a graphical user interface instead of the current CLI. Dynamic model selection is vital for high-quality estimation. Figure 10, which uses the same input data as used in Fig. 7, highlights the risks of incorrect model selection. Selecting an unfit model variogram leads to poor kriging outcomes, in this case leading to an entirely unreliable estimate of the distribution of $\mathrm{PM}_{2.5}$ in the mine.

Real-time airborne contamination modeling and monitoring is feasible using SAM-UG. SAM-UG processed the entire four-hour dataset in 144 s during a static test in which SAM-UG had access to all four hours of data upfront to eliminate the one-minute data delay from regular API calls. Although SAM-UG functions as a real-time monitoring platform, a static test where all data is presented and analyzed sequentially without delay allows for a comprehensive test of the time required for SAM-UG to receive data, select a model, estimate, and visualize. This required approximately 0.58 s to read the sensor data, select a model, and construct the visualization. Because SAM-UG constructs $9N$ covariance matrices during model selection for N installed sensors, the time complexity for model selection and distribution estimation can be approximated as $O(n)$. This means that the time required to estimate distributions scales approximately linearly with the number of fixed air quality monitors or the size of the estimation domain. Theoretically, this means that SAM-UG could manage upwards of 250 air quality sensors while remaining a suitable real-time monitoring platform. This would have the added benefit of supporting more sensors within the SAM-UG platform and allow coverage of significantly larger regions of underground mines. Further assessment of the SAM-UG platform will require additional testing to determine the practical upper limit to the number of installed sensors feasible for real-time monitoring, and care should be taken to select a reasonable number of sensors for installation. This number should ideally balance operational practicality for the monitoring application (underground mine) as well as the computational capacity of SAM-UG, data APIs and servers, and the host machine performing calculations and constructing estimates. Regardless, additional consideration should be given to the use of parallel computing to construct multiple covariance matrices at the same time. There exist potential computational barriers associated with the inversion of larger covariance matrices. Singular value decomposition, which is required for Moore-Penrose pseudoinversion, displaying a time complexity of $O(n^3)$.

As an initial attempt at a real-time contamination distribution monitoring platform, SAM-UG performs as expected.

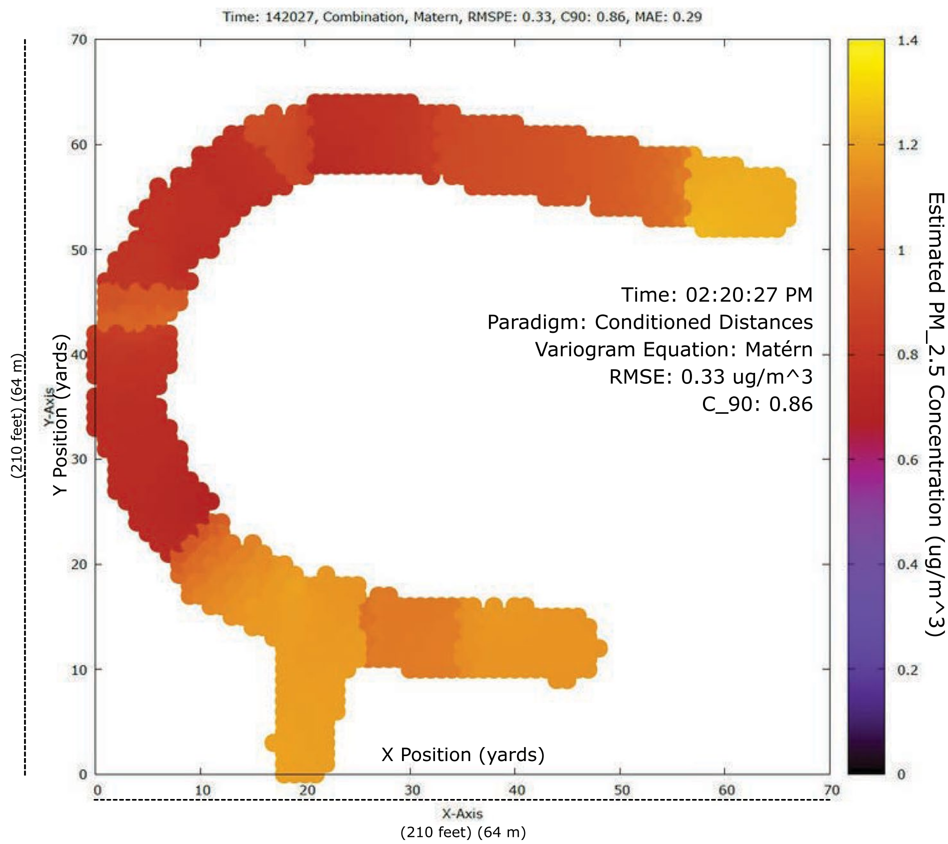

Fig. 9 Estimated contamination distribution at 2:20 PM using Matérn-Conditioned model. The plot as generated by SAM-UG has been altered in post-processing to provide clearer and more descriptive labels.

The limited, four-hour study period serves as an indication of a proof of concept for the platform. While this short timeframe does not fully capture the dynamic contamination conditions within mine ventilation systems as a result of changing ventilation conditions, generation rates, and seasonal variation, the work presented here indicates that SAM-UG may prove a viable avenue for real-time contamination monitoring beyond approaches available at current. Other monitoring methods like those employed by modern AMSs remain incredibly information-poor even while data-dense. By displaying contamination data as a distribution rather than individual points across a mine's footprint, SAM-UG provides a more holistic view of airborne contamination behavior by providing a reasonable estimate of contamination concentration between sensors. This is not otherwise feasible using mine ventilation network simulation, nor feasible in real-time or over large scales using CFD. SAM-UG provides a new avenue to contamination distribution monitoring using spatial statistics, and with performance somewhat comparable to the MSHA-approved Thermo Scientific PDM3700, which has a measurement uncertainty of 25% (Volkwein et al. 2006). Further assessment of the SAM-UG platform is required to fully validate the reliability of the system as well as determine next steps for reduction of total estimation uncertainty below the current 30.4%.

Because SAM-UG automatically selects the best performing statistical modeling method using REMLengine 1.7, no knowledge of spatial statistics or mine ventilation is required for its use in underground mines. As a statistical method, SAM-UG should also theoretically be species-agnostic, meaning its use can be transferred to other types of airborne contamination. Capable of functioning in real-time, SAM-UG may be a good candidate for underground mining operations wishing to monitor airborne contamination with higher spatial resolution than is otherwise feasible. Alongside Flowstar and EPAQS, SAM-UG forms an operations-ready software ecosystem and airborne contamination distribution monitoring platform aimed at real-time statistical air monitoring with minimal human intervention required. This would fundamentally change current airborne contamination monitoring practices in underground mines, providing operations with a clear, real-time picture of contamination behavior without then need of modeling using MVNs or CFD.

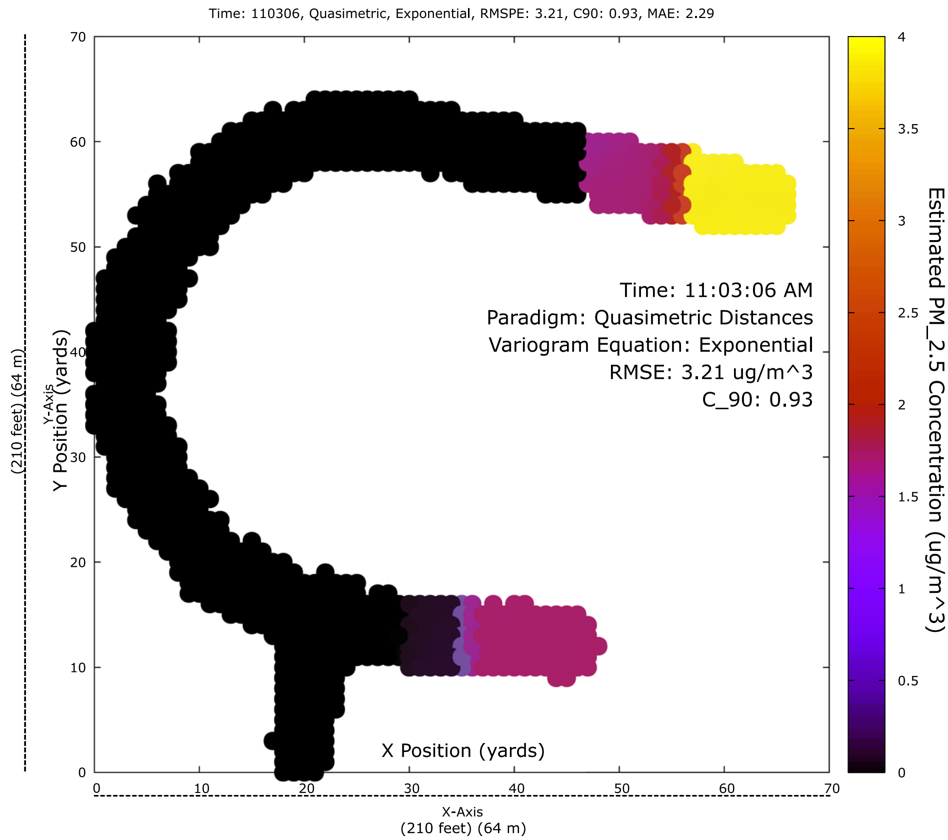

Fig. 10 Estimated contamination distribution at 11:03 AM using Exponential-Quasimetric model. Selection of an unfit modeling approach yields estimates which are unrealistic, invalid, and unreliable. The plot as generated by SAM-UG has been altered in post-processing to provide clearer and more descriptive labels

4 Limitations and future work

This work was limited to a single mine layout with a relatively simple ventilation regime. Kriging of air quality with simulated mine layouts has been conducted with similar performance (Brown Requist and Momayez 2024a). More work is needed for experimental validation of SAM-UG using more complex mine layouts. This work should also prioritize the development of external validation approaches including the use of computational fluid dynamics, especially the use of Euler-Lagrange models to demonstrate the physical validity of the resulting estimators constructed via kriging. Additionally, the number of sensors used and precision of sensors plays a major role in monitoring performance. Too few sensors will not provide sufficient model resolution, and we suggest preliminary assessment of data covariance be conducted using dynamic simulation in network analysis software such as Ventsim to arrive at an indication of reasonable ranges. Ideally, sensors should be placed within at least 2/3 the total distance of the Euclidean variogram range to provide sufficient coverage for reliable kriging. Low-precision sensors will of course increase estimation uncertainty. This application using low-cost particulate matter monitors is promising, but more precise sensors would have a measurable impact on the uncertainty of kriged models. We suggest that care should be taken during sensor selection to ensure that sensor uncertainty is below $\pm 15\%$ with respect to true values observed within the mine. Sensors with high limits of detection should be avoided at all costs, and selection of equipment must be operationally relevant. In essence, selected sensors should have measurement ranges and resolutions that are appropriate for in-mine conditions.

More work is needed to assess avenues for improved estimation performance. The inclusion of additional covariates such as temperature and relative humidity may help to explain observed spatial variability of $\mathrm{PM}_{2.5}$ mass concentrations (Tian et al. 2014; Zender-Świercz et al. 2024). Additionally, the authors are considering further validation of the SAM-UG platform by conducting a sensitivity analysis based on longer data collection time periods and an assessment of how each model responds to changes in input parameters such as sensor placement, number of sensors employed, data quality, and data stationarity under various spatial paradigms. This will provide an improved understanding of modeling behaviors, which will help to improve the estimation performance of the SAM-UG platform. Simply stated, estimation uncertainty with SAM-UG stands to be greatly improved. All observed data exhibit some level of measurement uncertainty, and this is no different in the case of the PurpleAirs. As a low-cost OPC monitor, the PurpleAir uncertainty is higher than more expensive particulate matter monitoring options. This is a reasonable operational trade-off for general area monitoring in underground mining environments, but the addition of geostatistical estimation uncertainty compounds this uncertainty. Further work will consider additional techniques to improve the numerical stability of kriging over quasimetric and otherwise non-stationary estimation domains, which may help to additionally reduce estimation uncertainty while providing a physics-based spatial basis for modeling.

SAM-UG is intended to serve as a species-agnostic contamination distribution monitoring platform. While it has been successfully tested using $\mathrm{PM}_{2.5}$ data, considerations should be made for the transferability of SAM-UG to other airborne contamination species such as carbon monoxide in underground coal operations as supplements to existing AMSs, elemental carbon as a proxy for diesel particulate matter assessment, and other mine gasses like NO_x and SO_x. This is especially pertinent given limitations to OPC technology which prevents its use to determine concentrations of specific particulate types, such as RCS and DPM. Considerations must be made, however, to the different transport phenomena governing gases compared to particulate matter. Future work will include these considerations to ensure that this modeling approach is indeed transferrable between particulate and gases. Full-scale validation work will be performed in conjunction with a large underground mining project across a variety of airborne contaminants, including NO_x, SO_x and $\mathrm{PM}_{2.5}$ with significantly longer study periods, at least 48 h. Additional investigation using real-time monitoring technology which can discern specific contaminant species, like what is currently possible across a range of gasses and elemental carbon will help to solidify the reliability of the SAM-UG platform outside of the large umbrella of $\mathrm{PM}_{2.5}$.

5 Conclusion

SAM-UG is a real-time contamination distribution modeling and monitoring platform that has been successfully trialed using low-cost OPC monitors for the assessment of $\mathrm{PM}_{2.5}$ mass concentration in a portion of an underground mining research facility in Sahuarita, Arizona. Supported by Flowstar and EPAQS, SAM-UG is part of a statistics-based air quality monitoring software ecosystem designed specifically for underground mining operations. With the use of spatial statistics, SAM-UG provides more spatial resolution than is possible with current AMSs or mine ventilation network simulation, and can model airborne contamination distributions in real-time, unlike CFD methods. This platform is intended to supplement existing contamination monitoring frameworks by providing a lightweight, computationally efficient, and reliable method to estimate and visualize airborne contamination distributions in real time. Its statistics-based design allows for portability across different types of airborne contamination while remaining adaptable and accurate across complex underground mining environments.

Developed in C++, SAM-UG uses LOOCV to consider nine different kriging approaches and automatically selects the best approach based on RMSE, the presence of negative kriging variances, and coverage of the 90% confidence interval by way of REMLengine 1.7. Over a four-hour study period, SAM-UG displayed an average estimation uncertainty of 27.9% and an RMSE of 0.54 $\mu g/m^3$. By dynamically selecting a kriging approach across three variogram equations (exponential, spherical, Matérn) and three spatial paradigms (linear, quasimetric, conditioned), SAM-UG provides the most reliable estimate of contamination distribution feasible without requiring human intervention or a robust understanding of spatial statistics. Although more work is needed to test SAM-UG for use with other airborne contamination species and to reduce estimation uncertainty, SAM-UG is a functional, real-time airborne contamination distribution modeling and monitoring platform that could present an alternative approach to contamination monitoring in underground mines.

The most recent versions of SAM-UG, EPAQS and Flowstar are available as Windows programs at no cost upon reasonable request. REMLengine 1.7 is available at no cost upon reasonable request for research purposes only. All data presented in this manuscript is available upon request.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00477-026-03192-4.

Author Contributions KBR: Conceptualization, data curation, formal analysis, investigation, software, visualization, writing—original draft, writing—review & editing. MM: Resources, writing—review & editing.

Funding This research was supported by the National Institute for Occupational Safety and Health under award number U60OH012351.

Data Availability Data sets generated during the study are available from the corresponding author on reasonable request. Each software program/programming package is available upon reasonable request to the corresponding author. Original mine geometry and ventilation information for this project is restricted and not available outside of exceptional circumstances.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Amoah NA, Xu G, Kumar AR, Wang Y (2023) Calibration of low-cost particulate matter sensors for coal dust monitoring. Sci Total Environ 859:160336

Armstrong M (1992) Positive definiteness is not enough. Math Geol 24:135–143

Armstrong M, Diamond P (1984) Testing variograms for positive-definiteness. J Int Assoc Math Geol 16:407–421

Armstrong M, Jabin R (1981) Variogram models must be positive-definite. J Int Assoc Math Geol 13:455–459

Baron PA (2016) Measurement of fibers. Technical report, National Institute for Occupational Safety and Health. https://stacks.cdc.gov/view/cdc/151249

Beauchamp M, Fouquet C, Malherbe L (2017) Dealing with non-stationarity through explanatory variables in kriging-based air quality maps. Spat Stat 22:18–46

Beauchamp M, Malherbe L, Fouquet C, Létinois L, Tognet F (2018) A polynomial approximation of the traffic contributions for kriging-based interpolation of urban air quality model. Environ Model Softw 105:132–152

Boisvert J, Deutsch CV (2008) Kriging in the presence of IVA using Dijkstra's algorithm. Centre for Computational Geostatistics

Bourgault G (1997) Spatial declustering weights. Math Geol 29:277–290

Brown Requist K, Momayez M (2024) Minimum cost pathfinding algorithm for the determination of optimal paths under airflow constraints. Mining 4(2):429–446

Brown Requist K, Lutz E, Momayez M (2023) Leveraging air quality sensing for carbon monoxide transport modeling in underground coal mines. In: Proceedings of 2023 application of computers and operations research in the minerals industry

Brown Requist K, Momayez M (2024a) Towards site-wide air mixture monitoring: Lessons learned from geostatistics. In: Proceedings of the 12th international mine ventilation congress

Brown Requist K, Momayez M (2025) Application of spatial basis functions for the estimation of airborne contamination distributions in underground mining environments. In: Proceedings of the 20th North American mine ventilation symposium

Chang P, Xu G, Huang J (2020) Numerical study on DPM dispersion and distribution in an underground development face based on dynamic mesh. Int J Min Sci Technol 30(4):471–475

Conn AR, Scheinberg K, Toint PL (1997) Recent progress in unconstrained nonlinear optimization without derivatives. Math Program 79:397–414

Cressie N (1991) Statistics for spatial data. John Wiley & Sons, Hoboken

Cressie N, Lahiri SN (1996) Asymptotics for REML estimation of spatial covariance parameters. J Stat Plan Inference 50(3):327–341

Davis BJ, Curriero FC (2019) Development and evaluation of geostatistical methods for Non-Euclidean-based spatial covariance matrices. Math Geosci 51(6):767–791

Deutsch C (1989) Declus: a fortran 77 program for determining optimum spatial declustering weights. Comput Geosci 15(3):325–332

Donoghue AM (2004) Occupational health hazards in mining: an overview. Occup Med 54(5):283–289

Duan J, Zhou G, Yang Y, Jing B, Hu S (2021) CFD numerical simulation on diffusion and distribution of diesel exhaust particulates in coal mine heading face. Adv Powder Technol 32(10):3660–3671

Fouedjio F (2017) Second-order non-stationary modeling approaches for univariate geostatistical data. Stoch Env Res Risk Assess 31(8):1887–1906

Gardner B, Sullivan PJ, Lembo AJ Jr (2003) Predicting stream temperatures: geostatistical model comparison using alternative distance metrics. Can J Fish Aquat Sci 60(3):344–351

Ghamari M, Soltanpur C, Rangel P, Groves WA, Kecojevic V (2022) Laboratory and field evaluation of three low-cost particulate matter sensors. IET Wirel Sensor Syst 12(1):21–32

Goovaerts P (1997) Geostatistics for natural resources evaluation, vol 483. Oxford University Press, Oxford

Gordon CC, Blackwell CL, Bradtmiller B, Parham JL, Barrientos P, Paquette SP, Corner BD, Carson JM, Venezia JC, Rockwell BM et al (2014) 2012 anthropometric survey of us army personnel: methods and summary statistics. Tech. Rep, Army Natick Soldier Research Development and Engineering Center MA

Halterman A, Sousan S, Peters TM (2018) Comparison of respirable mass concentrations measured by a personal dust monitor and a personal dataram to gravimetric measurements. Ann Work Expos Health 62(1):62–71

Heyde C (1994) A quasi-likelihood approach to the REML estimating equations. Stat Probab Lett 21(5):381–384

Kraft D, Schnepper K (1989) Slsqp-a nonlinear programming method with quadratic programming subproblems. DLR, oberpfaffenhofen 545

Laney AS, Weissman DN (2014) Respiratory diseases caused by coal mine dust. J Occup Environ Med 56:18–22

Lawson S, Woodgate M, Steijl R, Barakos G (2012) High performance computing for challenging problems in computational fluid dynamics. Prog Aerosp Sci 52:19–29

Little LS, Edwards D, Porter DE (1997) Kriging in estuaries: as the crow flies, or as the fish swims? J Exp Mar Biol Ecol 213(1):1–11

Marchant BP, Lark RM (2007) The matérn variogram model: implications for uncertainty propagation and sampling in geostatistical surveys. Geoderma 140(4):337–345

Marchant B, Lark R (2007) Robust estimation of the variogram by residual maximum likelihood. Geoderma 140(1–2):62–72

Martin R, Machuca-Mory D, Leuangthong O, Boisvert JB (2019) Non-stationary geostatistical modeling: a case study comparing lva estimation frameworks. Nat Resour Res 28(2):291–307

Mora L, Gadgil A, Wurtz E, Inard C (2002) Comparing zonal and CFD model predictions of indoor airflows under mixed convection conditions to experimental data. In: Proceedings of 3rd european conference on energy performance and indoor climate in buildings, Lyon, France, pp. 23–26

Oliver DS (1995) Moving averages for Gaussian simulation in two and three dimensions. Math Geol 27:939–960

Pyrcz MJ, Deutsch CV (2006) Semivariogram models based on geometric offsets. Math Geol 38(4):475–488

Rathbun SL (1998) Spatial modelling in irregularly shaped regions: kriging estuaries. Environmetr Off J Int Environmetr Soc 9(2):109–129

Requist KB, Momayez M (2025) An algorithm for the efficient placement of air quality sensors in underground mines. Min Metall Explor. https://doi.org/10.1007/s42461-025-01241-0

Requist K, Lutz E, Momayez M (2024) Near real-time interpolative algorithm for modelling air quality in underground mines. J S Afr Inst Min Metall 124(3):153–162

Rowland J III, Harteis S, Yuan L (2018) A survey of atmospheric monitoring systems in us underground coal mines. Min Eng 70(2):37

Sahu SK, Mardia KV (2005) A Bayesian kriged Lalman model for short-term forecasting of air pollution levels. J R Stat Soc: Ser C: Appl Stat 54(1):223–244

Shriwas M, Pritchard C (2020) Ventilation monitoring and control in mines. Min Metall Explor 37(4):1015–1021
Son P (1982) Embedding quasi-metric spaces in hilbert spaces. Aequationes Math 24:39–46
Tian G, Qiao Z, Xu X (2014) Characteristics of particulate matter (pm10) and its relationship with meteorological factors during 2001–2012 in beijing. Environ Pollut 192:266–274
Varouchakis EA, Koltsidopoulou MD, Pavlides A (2025) Designing robust covariance models for geostatistical applications. Stochastic Environ Res Risk Assess. https://doi.org/10.1007/s00477-025-02982-6
Ver Hoef JM (2018) Kriging models for linear networks and Non-Euclidean distances: cautions and solutions. Methods Ecol Evol 9(6):1600–1613
Vicedo-Cabrera AM, Biggeri A, Grisotto L, Barbone F, Catelan D (2013) A Bayesian kriging model for estimating residential exposure to air pollution of children living in a high-risk area in italy. Geospat Health 8(1):87–95
Volkwein JC, Vinson RP, Page SJ, McWilliams LJ, Joy GJ, Mischler SE, Tuchman DP (2006) Laboratory and field performance of a continuously measuring personal respirable dust monitor
Webster R, Oliver MA (2007) Geostatistics for environmental scientists. John Wiley & Sons, Hoboken
Wilson WA (1931) On quasi-metric spaces. Am J Math 53(3):675–684
Wolfe C, Cauda E, Yekich M, Patts J (2024) Real-time dust monitoring in occupational environments: a case study on using low-cost dust monitors for enhanced data collection and analysis. Min Metall Explor. https://doi.org/10.1007/s42461-024-01039-6
Xu R (2015) Light scattering: a review of particle characterization applications. Particuology 18:11–21
Yuan L, Zhou L, Smith AC (2016) Modeling carbon monoxide spread in underground mine fires. Appl Therm Eng 100:1319–1326
Zabihi M, Brinkerhoff J, Li R (2025) On Euler-Lagrange urans turbulence models for predicting the transient dispersion of aerosols indoors. Build Environ 271:112601
Zender-Świercz E, Galiszewska B, Telejko M, Starzomska M (2024) The effect of temperature and humidity of air on the concentration. Atmos Res 312:107733
Zhang H, Fava L, Cai M, Vayenas N, Acuña E (2021) A hybrid methodology for investigating DPM concentration distribution in underground mines. Tunn Undergr Space Technol 115:104042
Zheng S, Zhai ZJ, Wang Y, Xue Y, Duanmu L, Liu W (2022) Evaluation and comparison of various fast fluid dynamics modeling methods for predicting airflow around buildings. In: Building simulation, pp. 1–13 . Springer
Zhou L, Goodman G, Martikainen A (2013) Computational fluid dynamics (CFD) investigation of impacts of an obstruction on airflow in underground mines. Trans Soc Min Metall ExploratInc 332:505