Tuesday, September 23, 2025

From Trade Winds to Vog: Short-Term $PM_{2.5}$ Forecasting in Hilo

Literature Review for I-492 Project

by

Ailene Johnston

Advisor
Dr. Sridhar Ramachandran

Table of Contents

I. Next-hour $PM_{2.5}$, Forecasting for hilo

Air quality is inseparable from public health, and few pollutants carry the same weight as $PM_{2.5}$ (particulate matter ≤ 2.5 µm). Because these particles can slip deep into the lungs and bloodstream, they shape not only long-term health outcomes but also everyday choices. A sudden change in pollution levels can alter whether children are sent outdoors at school, workers are assigned to indoor or outdoor tasks, or families decide to open their windows. For this reason, short-lead forecasts - predictions on the scale of the next hour - are especially valuable, yet they remain underdeveloped compared to the daily or multi-day forecasts that dominate most studies.

In Hawaiʻi, the need for such forecasts is particularly acute. Residents face the usual mix of anthropogenic and meteorological pollution drivers, but they also live with volcanic smog, or "vog," produced by emissions from Kīlauea. This distinct setting poses challenges for both monitoring and forecasting. Low-cost sensors such as PurpleAir, widely adopted elsewhere, tend to overestimate $PM_{2.5}$ under humid conditions, prompting a series of correction models ranging from simple regressions to nationwide formulas and deep learning systems (Mathieu-Campbell et al., 2024; Barkjohn et al., 2021; Ghahremanloo et al., 2025). Meteorological research confirms the central role of boundary layer dynamics, wind, and humidity (Allabakash & Lim, 2020; Wang et al., 2019, 2023), while Hawaiʻi-specific studies point to synoptic disturbances such as Kona lows that move vog across islands (Tofte et al., 2017). Volcanic research has established both the chemical pathways of sulfate aerosol formation (Tam et al., 2016; Tang et al., 2020) and the health burdens of exposure (Brook et al., 2019). Finally, advances in forecasting methods, from autoregressive models to hybrid deep learning systems, continue to push accuracy forward, though questions about interpretability and transferability remain (Kristiani et al., 2022; Feng et al., 2024; Wu et al., 2025).

Despite this progress, Hawaiʻi has been largely absent from calibration and forecasting studies. Correction models are tuned to continental climates, forecasting horizons tend to favor daily averages rather than immediate conditions, and volcanic activity is often treated as an isolated hazard rather than an integrated input. This project responds to that gap by testing whether simple, interpretable models that incorporate recent $PM_{2.5}$ history, basic meteorological predictors, and volcanic activity flags can outperform persistence baselines in next-hour forecasts for Hilo, Hawaiʻi. In doing so, it aims to provide a tool that is both locally relevant and a broader test of how forecasting methods perform in underrepresented maritime and volcanic climates.

The past decade of research has pushed the science of $PM_{2.5}$ forward in several directions. Low-cost sensors such as PurpleAir have made dense monitoring networks possible, but their accuracy has been repeatedly challenged by environmental factors. Studies show that humidity in particular leads to systematic overestimation of particle concentrations. Researchers have responded with calibration approaches that range from simple regression adjustments to nationwide correction models now embedded in EPA's Fire and Smoke Map, and even deep learning systems that optimize accuracy across thousands of sites (Mathieu-Campbell et al., 2024; Barkjohn et al., 2021; Ghahremanloo et al., 2025).

Meteorological research has clarified how weather controls $PM_{2.5}$ far beyond emission rates alone. Boundary layer depth, wind strength and direction, and relative humidity all play decisive roles in whether particles disperse or stagnate (Allabakash & Lim, 2020; Wang et al., 2019, 2023). Work in Hawaiʻi adds another layer: synoptic disturbances such as Kona lows can transport vog across island chains, making meteorology as critical as emissions in determining exposure (Tofte et al., 2017).

Volcanic emissions themselves form a distinctive branch of this literature. Studies have tracked the chemical transformation of sulfur dioxide into sulfate aerosols, noting how sunlight and humidity accelerate this process (Tam et al., 2016; Tang et al., 2020). Health research ties vog episodes to asthma exacerbations, emergency visits, and cardiovascular strain, particularly in vulnerable populations (Brook et al., 2019). Efforts like the Vog Measurement and Prediction project show that plume forecasting is possible, though they also highlight the difficulty of capturing fine-scale variation under variable winds (Businger et al., 2015).

Forecasting methods as a whole are in transition. Statistical and persistence models remain important for their clarity, but they often lag when conditions change rapidly. Machine learning, especially LSTM and CNN architectures, has pushed accuracy forward by handling nonlinear relationships and temporal dependencies (Kristiani et al., 2022; Wu et al., 2025). Hybrid and ensemble methods now attempt to strike a balance by combining physical dispersion models with machine learning corrections to improve both resilience and transferability (Feng et al., 2024; Petrić et al., 2024; Özüpak et al., 2025).

Even with this progress, Hawaiʻi has been largely left out. Calibration models have not been tested in maritime air dominated by sea salt and constant humidity. Most forecasting frameworks target daily or multi-day horizons, not the next hour that matters for immediate decisions. And while volcanic emissions are studied as isolated hazards, they are rarely integrated as structured inputs into predictive models.

This project takes up that challenge. It develops a next-hour forecasting framework for Hilo, Hawaiʻi, anchored on regulatory-grade AQS monitors and supplemented with PurpleAir data where possible. By using simple but physically relevant predictors: recent $PM_{2.5}$ levels, wind direction and speed, and an indicator of volcanic activity - the study tests whether interpretable models such as logistic regression and shallow decision trees can outperform persistence baselines. The goal is to produce forecasts that are both credible to regulators and practical for communities.

While sensor calibration and $PM_{2.5}$ forecasting have advanced globally, Hawaiʻi remains an untested case. By integrating meteorology and volcanic activity into next-hour prediction, this project aims to close that gap, advancing local air quality management while also extending theory into a maritime volcanic climate.

II. Literature Review

2.1 Low-Cost Sensors and Calibration Challenges

PurpleAir and other low-cost sensors have changed the way $PM_{2.5}$ is monitored, offering high-density coverage that regulatory networks alone could never achieve. Their popularity in community science and state monitoring programs reflects both their affordability and accessibility. Yet their readings are far from perfect. Raw outputs often drift under certain weather conditions, raising questions about how to interpret them and what kinds of corrections are needed. The research in this area can be grouped into three threads: environmental biases, calibration methods, and quality assurance practices.

K. K. Barkjohn et al.: Development and application of a United States-wide correction for $PM_{2.5}$ data **4623**

Figure 2. State, local, and tribal (SLT) air monitoring sites with collocated PurpleAir sensors, including regions used for correction model evaluation.

2.1.1 Biases and Environmental Influences

Mathieu-Campbell et al. (2024) ran a year-long collocation study in the southeastern U.S., where humidity is persistently high. By pairing PurpleAir units with FEM monitors, they found that raw sensor readings inflated $PM_{2.5}$ levels during humid periods, sometimes by 30–40%. Their seasonal analysis showed the problem was not uniform: summer peaks were consistently the worst, pointing directly to hygroscopic particle growth as the culprit.

6742 **M. E. Mathieu-Campbell et al.: Calibration of PurpleAir low-cost PM sensors**

Table 2. Semi-supervised clustering model development (model fit with hourly data) and application of the hourly model to daily data. Temperature is in units of degrees Celsius.

Parameters		Model fit with hourly data				Model fit to daily data			
Clusters (number of observations)	Models	R^2 (%)	RMSE ($\mu g\,m^{-3}$)	MAE ($\mu g\,m^{-3}$)	R (%)	R^2 (%)	RMSE ($\mu g\,m^{-3}$)	MAE ($\mu g\,m^{-3}$)	R (%)
RH $\leq$ 50 (59405)	$2.738732 + 0.425834\,PA_i - 0.008944\,RH_i + 0.079210\,T_i$	71	2.96	1.86	84	88	2.04	1.46	94
RH > 50 (100243)	$7.230374 + 0.412683\,PA_i - 0.085278\,RH_i + 0.070655\,T_i$	74	2.92	2.02	86	73	2.33	1.68	85

Barkjohn et al. (2021) expanded the scope by pulling together nearly 12,000 collocated daily averages from 16 states. Their study confirmed a national pattern - raw PurpleAir data overstated $PM_{2.5}$ by about 40% on average. Errors were largest during wildfire smoke and high-humidity episodes. Their work made clear that while the bias is widespread, its size and character vary across regions and seasons.

Ghahremanloo et al. (2025) examined over 1,500 sites, again finding that humidity and temperature consistently drove overestimation. Their contribution was methodological: they trained deep learning models that captured nonlinear relationships among these variables. Accuracy improved, but the cost was interpretability - black-box models explained little about why corrections worked, only that they did.

Taken together, these studies leave no doubt that humidity is the main source of bias, with temperature playing a supporting role. Each paper contributes differently: Mathieu-Campbell shows how the bias unfolds locally, Barkjohn establishes its consistency nationwide, and Ghahremanloo demonstrates how modern AI can correct it, even if the inner workings remain opaque.

4628 K. K. Barkjohn et al.: Development and application of a United States-wide correction for $PM_{2.5}$ data

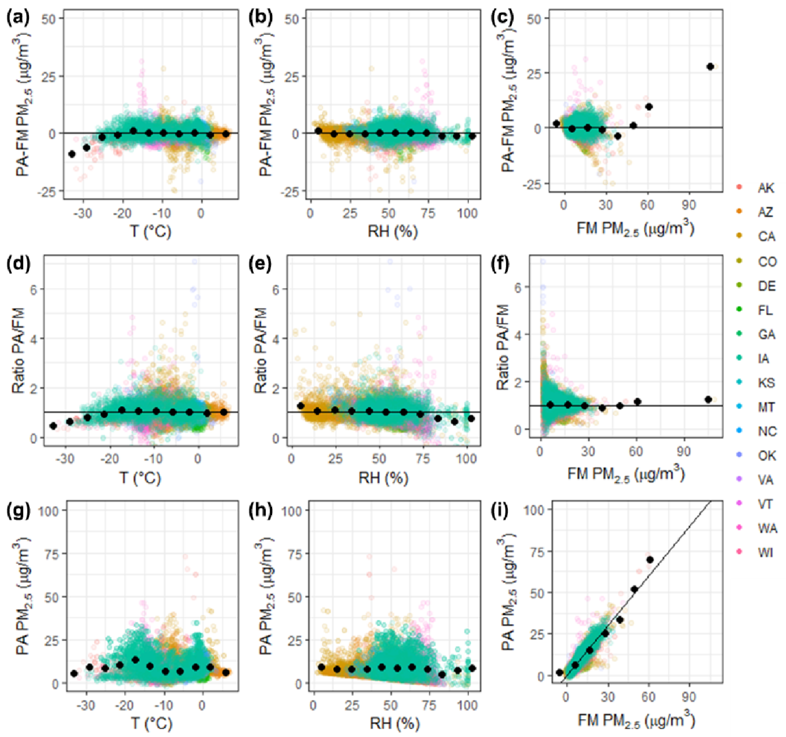


Figure 5. Error and ratio between corrected PurpleAir (PA) and FRM or FEM measurements are shown along with corrected PurpleAir $PM_{2.5}$ data (corrected using Eq. 10) as influenced by temperature, RH, and FRM or FEM $PM_{2.5}$ concentration. Colors indicate states, and black points indicate averages in 10 bins.

2.1.2 Calibration and Correction Models

Once bias was established, the natural next step was correction. Mathieu-Campbell's regression model, tuned to humidity, provided a straightforward way to bring PurpleAir readings closer to FEM monitors in a specific climate zone. It's simple, transparent, and effective on a local scale.

M. E. Mathieu-Campbell et al.: Calibration of PurpleAir low-cost PM sensors 6741

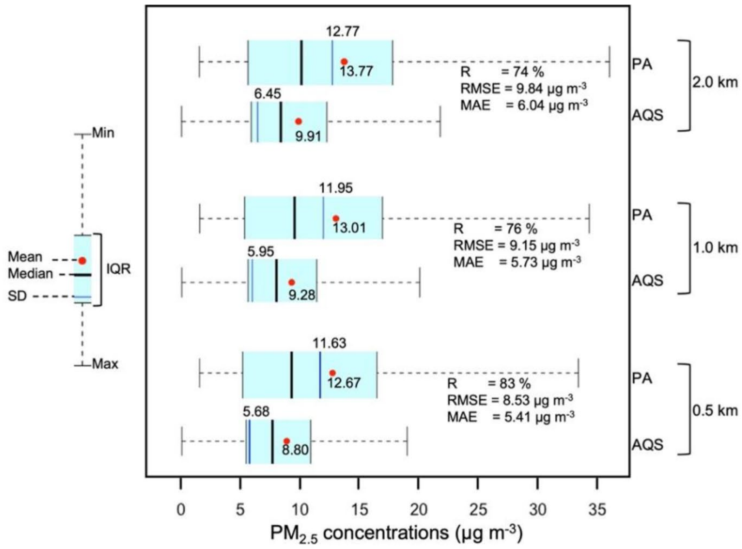
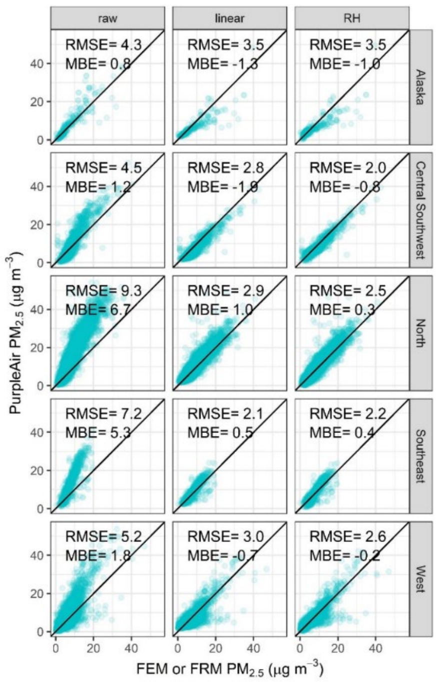


Figure 3. Descriptive and error metrics for C_{AQS} and raw C_{PA} for PurpleAir sensors within a 0.5, 1.0, and 2.0 km radii of each FRM or FEM monitor.

Barkjohn's team, by contrast, went for generalizability. They built a U.S.-wide equation that balanced accuracy with ease of use, and the EPA adopted it directly into the Fire and Smoke Map. This move marked a turning point, one of the few times a peer-reviewed calibration approach has been operationalized nationally.



Ghahremanloo's neural network approach offered a leap in predictive performance, but at the expense of transparency. Their model captured nonlinearities better than linear regression, but it left users with less insight into why corrections were necessary. For policy or regulatory contexts, that trade-off is not trivial.

Together, these studies trace a progression from simple, local corrections to nationwide formulas and finally to high-capacity machine learning. Each step improves accuracy but also shifts the balance between interpretability and operational practicality. Calibration methods fall on a spectrum. At one end are simple, easily explained regression models; in the middle are national formulas that work well across settings and can be implemented at scale; at the other are complex neural networks that maximize accuracy but shed interpretability.

2.1.3 QA/QC and Reference Comparisons

Calibration is only part of the story. Without quality control, even the best equations fall short.

The EPA Sensor Loan Program QAPP (2024) codified practices that many researchers had already been using: collocating PurpleAir units with reference monitors, checking consistency between their dual channels, and applying performance thresholds. These protocols were designed to weed out unreliable data and standardize outputs before applying corrections.

Barkjohn's national study emphasized the importance of collocation. They showed that corrections work best when sensors have been evaluated side by side with FEM or FRM instruments, not in isolation. Earlier EPA guidance (2017) makes the same point: low-cost sensors can enrich monitoring networks, but they cannot replace reference methods.

The message from these sources is consistent: low-cost sensors are valuable only when paired with rigorous QA/QC. Their greatest strength is high-resolution coverage, but their readings need anchoring against reference-grade instruments.

2.1.4 Synthesis of Findings on Low-Cost Sensor Performance

The body of work on PurpleAir tells a balanced story. On one hand, these sensors offer unprecedented density and accessibility. On the other, their raw data cannot be trusted without correction. Humidity-driven bias dominates, but calibration models - whether regression-based, national formulas, or neural networks - can reduce error substantially. QA/QC ensures that these methods rest on solid ground.

The gap, though, is geographic. Nearly all of this research has been done in continental U.S. settings. Hawai'i's mix of trade winds, maritime aerosols, and volcanic emissions creates an atmospheric environment not represented in existing models. Whether the U.S.-wide correction or deep learning models trained on mainland data will transfer to Hawai'i is an open question, and it is exactly the kind of problem this project is designed to test.

2.2 Meteorological Drivers of $PM_{2.5}$ in Island Environments

The way particulate matter behaves in the atmosphere is inseparable from meteorology. Emissions may be constant, but the difference between a clean day and a hazardous day often comes down to whether the atmosphere is dispersing or trapping pollutants. Boundary layer dynamics, temperature inversions, wind, humidity, and synoptic-scale weather patterns all play a role in shaping $PM_{2.5}$ concentrations at the surface. The studies reviewed here explore these relationships at multiple scales, from localized observations of boundary layer height (BLH) to regional analyses of weather systems that carry volcanic emissions across Hawai'i.

2.2.1 Boundary-Layer Height and Inversions

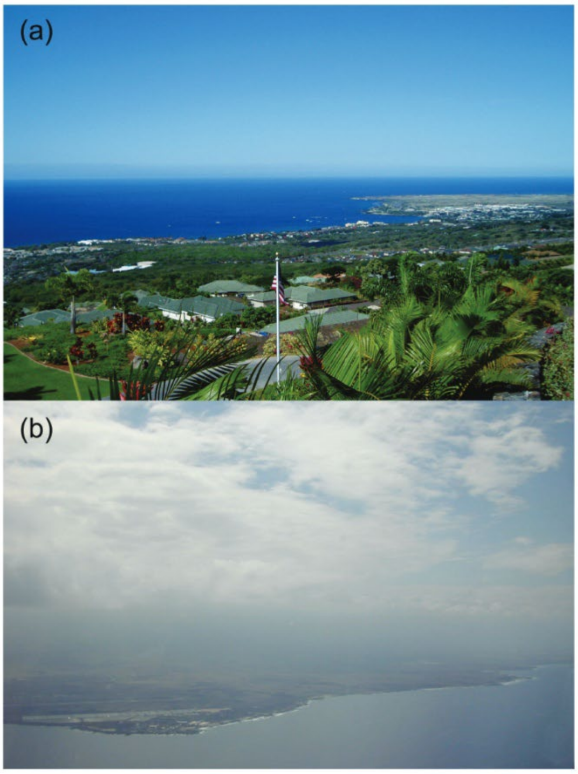


FIG. 3. (a) Kona coast on a clear day. (b) Vog layer trapped beneath trade wind inversion along the Kona coast.

The height of the planetary boundary layer (BLH) is one of the most important factors for surface $PM_{2.5}$. A shallow BLH traps pollutants close to the ground, while a deep BLH promotes vertical mixing and dilution.

Allabakash and Lim (2020) used radiosonde and reanalysis data to characterize BLH climatology over Korea. They showed that shallow BLHs were most common in winter and strongly associated with pollution episodes. Their work emphasized that stability, vertical wind shear, and subsidence control BLH variability, making these meteorological conditions indirect but powerful drivers of surface $PM_{2.5}$. They also highlighted diurnal variation: boundary layers deepened during the day with heating, diluting pollutants, and collapsed at night, often resulting in surface build-up.

Wang et al. (2019) used a combination of lidar aerosol profiling and turbulence measurements in Beijing to directly connect BLH with $PM_{2.5}$ levels. They observed sharp negative correlations: concentrations rose rapidly as BLH dropped, sometimes doubling during nighttime inversion events. They also quantified how turbulent kinetic energy modified dispersion; BLH dynamics are not static but evolve on hourly timescales.

Shallow BLHs act like a lid on pollution. BLH must be treated as a central predictor in short-term forecasts, since even modest changes in vertical mixing can mean the difference between moderate and severe pollution.

2.2.2 Local Meteorological Variables

While boundary layer height governs vertical dispersion, day-to-day variability in $PM_{2.5}$ is also strongly tied to local meteorological factors. Wind speed and direction, temperature, humidity, and precipitation all interact to influence how particles accumulate or disperse. Studies across Asia and Latin America illustrate how these variables work in combination to shape surface air quality.

Wang et al. (2023) conducted a scenario-based perturbation experiment in 14 of China's most polluted cities, including Handan, Shijiazhuang, and Baoding. Using the WRF-CMAQ model, they systematically adjusted five variables - planetary boundary layer height (PBLH), wind speed (WS), temperature (T), water vapor mixing ratio (Q), and precipitation (PCP)—to measure their impact on simulated $PM_{2.5}$. They ran 21 different scenarios by perturbing each variable ±10–20% from baseline January 2017 conditions. Wind speed, PBLH, and precipitation had the strongest linear effects. For example, a 10–20% decrease in wind speed led to a significant rise in $PM_{2.5}$ concentrations across all study cities, while increases in wind or precipitation reduced pollution.

Relative humidity (proxied through Q) and temperature effects were more complex: under some conditions, higher humidity and temperature lowered $PM_{2.5}$, but in other cities, they contributed to increases. Their model validation showed correlation coefficients (R) between 0.41–0.74 across cities, confirming reasonable predictive skill.

Gutiérrez-Avila et al. (2022) focused on Mexico City and surrounding areas, applying machine learning and statistical methods to predict both daily mean and one-hour maximum $PM_{2.5}$. They combined ground-based monitors with meteorological inputs such as temperature, humidity, and wind. Their work added the value of meteorological predictors: when included, model performance improved significantly, particularly for short-term peaks. High relative humidity consistently amplified particle size due to hygroscopic growth, while temperature shifts influenced secondary aerosol formation. Integrating these variables allowed forecasts to capture pollution spikes that would otherwise be missed if only emissions data were used.

Local meteorology cannot be ignored in $PM_{2.5}$ forecasting. Wang et al. demonstrated through controlled model perturbations that wind speed and PBLH are the most consistent drivers of pollution accumulation across Chinese megacities. Gutiérrez-Avila reinforced that in Mexico, temperature and humidity shape both the mass and chemistry of $PM_{2.5}$, directly improving model accuracy. Dispersion and chemical transformation are inseparable from local weather conditions, making them essential predictors for forecasting models.

2.2.3 Large-Scale Weather Patterns

While boundary layer height and local meteorological variables explain much of the day-to-day variation in urban $PM_{2.5}$, larger-scale circulation patterns determine how far pollution travels and which populations are exposed. This is especially important in Hawaiʻi, where emissions from Kīlauea Volcano can affect communities across multiple islands depending on the prevailing synoptic setup.

Tofte et al. (2017) analyzed 101 vog episodes on Oʻahu between 2005 and 2015, combining NCEP reanalysis data with hourly $PM_{2.5}$ observations. They used synoptic classification techniques to identify the weather patterns most favorable for vog transport. Three setups dominated: (1) pre-cold fronts, (2) upper-level disturbances, and (3) Kona lows.

Pre-cold fronts were most significant, often lasting up to four days and producing sustained high $PM_{2.5}$ episodes across Oʻahu. Upper-level disturbances and Kona lows also contributed, though typically for shorter periods. Their results highlighted that vog exposure in Honolulu and surrounding communities was not simply a function of emissions on Hawaiʻi Island but of synoptic conditions that altered regional airflow. By showing that specific weather patterns repeatedly coincided with high $PM_{2.5}$, Tofte et al. established a predictive link between synoptic meteorology and downwind exposure risk.

Tang et al. (2020) focused on the extreme 2018 Kīlauea eruption, when fissure vents released massive amounts of SO_2 and PM precursors. They used the WRF-Chem model to simulate dispersion and chemical transformation of volcanic emissions under different meteorological regimes. Their results showed that prevailing trade winds generally carried vog westward over the Pacific, sparing most of Hawaiʻi Island's population centers. However, when winds weakened or reversed during atypical synoptic conditions, sulfate aerosols blanketed communities across the Big Island and even extended to Maui and Oʻahu.

Tang et al. also documented how high humidity and solar radiation accelerated the conversion of SO_2 into sulfate aerosols, amplifying surface-level concentrations during stagnant conditions. Their simulations provided concrete evidence that the same emission source can produce very different exposure outcomes depending on the synoptic setup.

Hawaiʻi's air quality is tightly coupled to large-scale circulation patterns. Tofte et al. showed that repeatable synoptic regimes govern vog transport to Oʻahu, while Tang et al. quantified how these conditions influence both dispersion and chemical transformation during a major eruption. Local emissions alone cannot explain $PM_{2.5}$ exposure in Hawaiʻi; the broader meteorological context must be accounted for.

2.2.4 Synthesis of Findings on Meteorological Drivers

The research on meteorological drivers of $PM_{2.5}$ brings a consistent conclusion: pollution outcomes are often determined less by emission strength and more by the atmosphere's ability to disperse or trap particles. Across studies, three scales of influence can be observed.

At the boundary layer scale, Allabakash and Lim (2020) showed that shallow BLHs during winter in Korea frequently coincided with high $PM_{2.5}$, while Wang et al. (2019) provided direct observational evidence from Beijing that concentrations doubled as BLH collapsed overnight. BLH acts as a "lid," regulating vertical mixing and making it a critical input for any forecast model.

At the local meteorological scale, Wang et al. (2023) concludes through perturbation experiments that wind speed, PBLH, and precipitation exert the strongest linear control over $PM_{2.5}$ across 14 Chinese megacities. Even a 10–20% reduction in wind speed caused sharp increases in pollution, a finding echoed in Mexico by Gutiérrez-Avila et al. (2022), who showed that including wind, humidity, and temperature significantly improved forecast skill. Local weather variables are not optional "add-ons" but essential drivers of pollution dynamics.

At the regional or synoptic scale, Tofte et al. (2017) identified recurring weather patterns that repeatedly carried vog from Hawaiʻi Island to Oʻahu, while Tang et al. (2020) simulated how trade winds, humidity, and solar radiation shaped dispersion and chemical transformation during the 2018 eruption. Hawaiʻi's exposure cannot be understood from emissions alone; the broader circulation dictates who is affected and when.

Meteorology influences $PM_{2.5}$ at every scale: BLH sets the vertical boundary, local weather controls near-surface concentrations, and synoptic systems redistribute pollution across regions. For forecasting, meteorological predictors must be integrated from the outset, not treated as secondary. The majority of these studies are drawn from East Asia or continental contexts. Hawaiʻi's unique blend of persistent trade winds, frequent inversions, and volcanic emissions remains underexplored.

Local weather cannot be understood in isolation - regional circulation patterns ultimately shape whether pollution disperses or lingers, a dynamic that becomes especially important in island environments like Hawaiʻi.

2.3 Volcanic Episodes and Vog Impacts in Hawaiʻi

Volcanic eruptions and ongoing degassing events are a defining feature of Hawaiʻi's air quality. Emissions of sulfur dioxide (SO_2) and other gases from Kīlauea Volcano undergo chemical transformation in the atmosphere, producing fine sulfate aerosols that contribute heavily to $PM_{2.5}$. These particles, commonly referred to as "vog" (volcanic smog), are not only a public health concern but also a forecasting challenge, as their behavior depends on emission rate, atmospheric chemistry, and prevailing winds. The literature on volcanic episodes emphasizes three areas: (1) the chemical composition and transformation of volcanic emissions, (2) health impacts associated with vog exposure, and (3) modeling and forecasting of volcanic plumes.

2.3.1 Composition and Transformation of Volcanic Emissions

Understanding how volcanic gases become $PM_{2.5}$ is central to predicting vog episodes. Studies have focused on the conversion of SO_2 into sulfate aerosols under different meteorological conditions.

Tam et al. (2016) provided a detailed characterization of volcanic air pollution over Hawaiʻi. They documented how SO_2 emissions from Kīlauea, once released into the atmosphere, undergo oxidation to form sulfate aerosols. This process is accelerated by sunlight and high humidity, making vog events particularly severe during humid daytime conditions. Their study also described the typical composition of vog, which included sulfate particles, trace metals, and acidic components, all of which contribute to respiratory irritation.

Tang et al. (2020) expanded on this during the 2018 Kīlauea eruption using WRF-Chem simulations. They showed that conversion rates of SO_2 to sulfate depended heavily on meteorology. Under strong trade winds, emissions dispersed quickly and sulfate formation was limited. Under stagnant conditions, high humidity and strong solar radiation accelerated sulfate production, leading to hazardous surface-level concentrations. Their modeling results quantified how sulfate aerosols could persist downwind for days, making them a dominant contributor to $PM_{2.5}$ across the island chain.

Volcanic SO_2 is consistently transformed into fine sulfate aerosols, with humidity and solar radiation accelerating the process. Forecasting volcanic $PM_{2.5}$ therefore requires both emission data and real-time meteorological inputs to capture transformation rates.

2.3.2 Health Impacts of Vog

Beyond the atmospheric processes, vog poses well-documented health risks, particularly for cardiovascular and respiratory systems.

Brook et al. (2019) investigated the link between volcanic smog and cardiometabolic health in Hawai‘i, focusing on hypertension patients. Their study tracked daily blood pressure measurements and clinical markers across multiple vog events, showing that systolic pressure increased measurably on days with elevated $PM_{2.5}$ concentrations. Vog acts as a chronic stressor for individuals with pre-existing conditions, compounding the risks associated with long-term exposure to fine particulate matter.

Other health-focused studies cited by Tam et al. (2016) extended this evidence to the population level, documenting increases in emergency room visits during acute vog episodes, particularly for asthma exacerbations and other respiratory complications. Volcanic aerosols are not simply a nuisance in terms of visibility, but a tangible health hazard with measurable outcomes in both clinical and community settings.

Vog exposure is not only a scientific curiosity but a significant public health issue. Cardiovascular and respiratory impacts are most acute for vulnerable populations, making reliable forecasts urgently needed.

2.3.3 Modeling and Forecasting Volcanic Plumes

Given the episodic and often extreme nature of vog, modeling tools are essential for forecasting plume transport and concentrations.

Businger et al. (2015) described efforts to observe and forecast vog dispersion from Kīlauea using both observations and numerical models. They highlighted the Vog Measurement and Prediction (VMAP) project, which integrates real-time SO_2 emissions data with meteorological models to predict plume trajectories. Their evaluation showed that models captured broad transport patterns well but struggled with local detail, especially during periods of weak or variable winds. The study emphasized the importance of coupling volcanic emission data with high-resolution weather forecasts for operational predictions.

Tang et al. (2020) reinforced this by showing how WRF-Chem simulations during the 2018 eruption could reproduce plume transport across the Pacific. Their results also revealed that model skill depended strongly on accurate meteorological inputs, particularly wind fields and humidity.

Forecasting volcanic $PM_{2.5}$ requires models that integrate emission strength, chemistry, and meteorology. While current tools like VMAP provide valuable guidance, their limitations during variable wind conditions make a need for ongoing refinement.

2.3.4 Synthesis and Findings on Volcanic Episodes

Research on volcanic emissions and $PM_{2.5}$ makes clear that vog is both chemically distinct and meteorologically sensitive. Tam et al. (2016) and Tang et al. (2020) showed that sulfate aerosol formation is rapid under humid, sunny conditions, making transformation rates highly variable. Health studies such as Brook et al. (2019) demonstrated that these aerosols are not benign but have measurable cardiovascular and respiratory impacts. Modeling efforts by Businger et al. (2015) and Tang et al. (2020) established that forecasting vog is feasible but constrained by the accuracy of both emission inventories and meteorological inputs.

For Hawai‘i, volcanic activity is not a background factor but a primary driver of $PM_{2.5}$. Unlike continental regions where anthropogenic sources dominate, Hawai‘i's most severe $PM_{2.5}$ events stem from natural emissions transformed by weather. This makes the development of tailored, high-resolution volcanic forecasting systems critical for both research and public health protection.

2.4 Forecasting and Modeling Approaches

Forecasting $PM_{2.5}$ concentrations has advanced rapidly over the past decade, driven by new data sources and computational methods. Traditional statistical models provided the first predictive tools, but their limitations in capturing nonlinear dynamics opened the door to deep learning and ensemble techniques. More recently, hybrid models that combine physical understanding with machine learning have gained attention for their ability to balance interpretability and accuracy. The literature on forecasting $PM_{2.5}$ can be organized into three approaches: (1) statistical and time-series models, (2) deep learning methods, and (3) hybrid and ensemble approaches.

2.4.1 Statistical and Time-Series Models

Statistical and time-series models formed the foundation of $PM_{2.5}$ forecasting. Their main strength lies in their interpretability and relatively modest data requirements, but they often fall short when conditions change abruptly.

Equation (ARIMA(p,d,q)):

$$y_t = c + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t$$

Where p is the autoregressive order, d the differencing term, and q the moving average order. Here, ϕ_i are autoregressive coefficients, θ_j are moving average coefficients, and ϵ_t is the error term.

Wu et al. (2025) provided a detailed review of these approaches, including autoregressive integrated moving average (ARIMA), seasonal ARIMA (SARIMA), vector autoregression (VAR), and generalized autoregressive conditional heteroskedasticity (GARCH) models. In empirical applications across Asia and North America, ARIMA and SARIMA were able to capture broad seasonal and daily patterns, achieving correlations in the 0.6–0.8 range for day-ahead predictions. Yet, their performance deteriorated during rapid pollution spikes driven by shifts in weather, where forecast errors increased sharply. VAR models offered some improvement by capturing relationships between pollutants and meteorological drivers, though their reliance on complete datasets limited their operational use. GARCH models proved especially useful in characterizing volatility, allowing them to better anticipate short-term swings, but only when high-frequency data streams were available.

Statistical and time-series models provide transparency and solid baselines but fall short when faced with nonlinear or sudden changes. They are best suited as benchmarks or in combination with more flexible methods.

2.4.2 Deep Learning Models

Deep learning approaches have become the dominant research focus for $PM_{2.5}$ forecasting because of their ability to capture nonlinear dependencies among emissions, meteorology, and atmospheric chemistry.

Kristiani et al. (2022) applied Long Short-Term Memory (LSTM) networks to short-term $PM_{2.5}$ prediction. Using meteorological variables and historical pollution data as inputs, they demonstrated that LSTMs outperformed traditional statistical models in both RMSE and MAE. LSTMs can capture temporal dependencies across multiple timescales, making them particularly effective for multi-step forecasts.

Key LSTM update equations:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \quad \tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \quad h_t = o_t \odot \tanh(C_t)$$

Where f_t, i_t, o_t are the forget, input, and output gates, C_t is the memory cell, and h_t is the hidden state.

Bekkar et al. (2021) explored deep learning applications for air pollution prediction in smart city contexts. They tested multilayer perceptrons (MLPs) and convolutional neural networks (CNNs) alongside LSTMs, finding that performance varied by data type: CNNs excelled at handling spatial correlations, while LSTMs were superior for temporal sequences. Their work underscored the importance of aligning model architecture with data characteristics.

Wu et al. (2025) also evaluated recurrent neural networks (RNNs) and gated recurrent units (GRUs), emphasizing their strength in capturing sequential patterns while warning about overfitting in data-limited contexts.

Deep learning methods consistently outperform statistical baselines, especially for multi-step or nonlinear forecasts. However, they require large datasets and careful tuning, and their "black-box" nature makes them less interpretable for policy applications.

2.4.3 Hybrid and Ensemble Approaches

To address the weaknesses of single-model approaches, researchers have turned to hybrid and ensemble methods that combine statistical, machine learning, and physical modeling.

Feng et al. (2024) developed a hybrid system that linked atmospheric dispersion modeling with machine learning correction layers at a fine 1 km spatial resolution. Their results showed substantially lower RMSE compared to stand-alone machine learning models, especially in areas with strong local variability. Petrić et al. (2024) compared ensemble learning methods—including random forests and gradient boosting—to deep learning architectures like LSTMs across four urban regions. Random forest consistently delivered the lowest RMSE in three of the four sites, indicating that ensembles are more resilient when dealing with noisy or incomplete datasets. Özüpak et al. (2025) further confirmed this resilience by testing models across cities with contrasting climates: ensembles held steady across humid and arid conditions, while deep learning models showed larger swings in accuracy.

Example performance comparison (adapted from Feng et al. 2024; Petrić et al. 2024; Özüpak et al. 2025):

Model Type	RMSE (μg/m³)	MAE (μg/m³)	R^2	Notes
ARIMA (baseline)	12.8	10.5	0.52	Simple, interpretable
Random Forest	10.2	8.7	0.63	Captures nonlinearities
LSTM (deep learning)	8.9	7.4	0.71	High accuracy, low transparency
Hybrid (Dispersion + ML)	7.6	6.5	0.78	Best trade-off, still emerging

Saminathan and Malathy (2023) added to this body of work by showing that ensembles integrating meteorological variables such as wind speed and temperature improved classification of high versus low pollution days, proving that even relatively simple weather features can sharpen predictive skill.

Hybrid and ensemble approaches appear to offer the best of both worlds: the accuracy of machine learning with the stability and generalizability needed for operational use. By combining models, researchers can mitigate weaknesses of individual methods and adapt to diverse data environments.

2.4.4 Synthesis and Findings on Forecasting Approaches

The trajectory of $PM_{2.5}$ forecasting research mirrors the evolution of data science. Statistical models provided interpretable starting points but struggled with nonlinear dynamics. Deep learning brought major performance gains, especially for short-term and multi-step forecasts, but raised issues of interpretability and data demand. Hybrid and ensemble methods represent the latest wave, offering accuracy gains while also improving robustness across different contexts.

For Hawaiʻi, where meteorology and volcanic emissions interact in complex ways, hybrid or ensemble approaches are most promising. They can incorporate both physical models of volcanic plumes and machine learning layers trained on meteorological data, allowing forecasts to capture both chemistry and circulation. Building forecasting systems that merge physical insight with data-driven learning aligns closely with the unique challenges of predicting vog and $PM_{2.5}$ across the islands.

2.5 Critical Evaluation

The research on $PM_{2.5}$ monitoring and forecasting has grown rapidly, but it is uneven. Some areas are supported by rich empirical data and carefully validated models, while others are still speculative or regionally narrow. Collectively, the body of work illustrates the promise of low-cost sensing and advanced forecasting, yet it also shows how much depends on context and how fragile certain conclusions become once they leave the environments in which they were developed.

The studies on low-cost sensors highlight this tension clearly. Collocation experiments and calibration models have improved confidence in PurpleAir and similar devices, but their heavy reliance on relative humidity as a correction factor risks oversimplifying a more complex problem. The regression and empirical models tested by Mathieu-Campbell and Barkjohn demonstrate how practical and transferable these approaches can be, yet their limits are obvious: what works across the continental U.S. may not hold in places where particles are chemically distinct, as in Hawai‘i's volcanic environment. Ghahremanloo's deep learning approach pushes accuracy further, but it does so at the cost of transparency. In regulatory contexts, where decision makers must understand how a correction operates rather than simply trust the output, this loss of interpretability is more than a technical quibble - it determines whether the method is viable in practice.

Meteorological drivers reveal a comparable pattern. Boundary layer height and wind speed emerge repeatedly as consistent predictors of pollution buildup, with studies in China and Korea supporting this claim through large datasets and strong correlations. Work in Mexico shows how temperature and humidity interact with chemical processes to shape outcomes. While these findings are convincing, they also leave open questions about how to weight such variables in Hawai‘i, where persistent trade winds are punctuated by Kona lows and synoptic disturbances. The literature provides robust answers for continental megacities but only partial guidance for small, isolated islands.

The research on volcanic emissions is among the most directly relevant to Hawai‘i, yet it too remains incomplete. Studies such as Tam et al. and Tang et al. confirm that sulfate aerosols are the key pathway from SO_2 to $PM_{2.5}$ and quantify how humidity and radiation accelerate this transformation. Health research shows measurable cardiovascular and respiratory impacts during vog episodes. Forecasting systems, however, are still catching up. Businger's VMAP project is a rare attempt to design a model specifically for Hawai‘i, but it struggles with local detail when winds shift. In effect, the science has mapped the broad outlines of how vog forms and disperses but has yet to achieve operational accuracy at the neighborhood scale where exposure occurs.

The forecasting literature is perhaps the most fragmented of all. Statistical and time-series models remain attractive for their clarity and modest data requirements, yet they consistently fail to capture the nonlinear dynamics that dominate real-world pollution episodes. Deep learning methods, especially LSTMs and CNNs, achieve higher accuracy but demand large, clean datasets and careful tuning. Ensemble and hybrid models attempt to reconcile these strengths by combining physical insights with machine learning corrections, but most remain untested outside of research contexts. For Hawaiʻi, where volcanic emissions and meteorology interact in unusual ways, hybrid approaches appear promising, yet the literature has not demonstrated how they would perform under such conditions.

The studies reviewed here demonstrate the importance of calibration, the centrality of meteorology, the health burden of vog, and the potential of machine learning. At the same time, they reveal gaps in transferability, interpretability, and regional specificity. Much of the research has been conducted in continental settings where anthropogenic sources dominate, while Hawaiʻi's atmosphere is shaped by a very different mix of trade winds, marine aerosols, and volcanic emissions. Bridging this gap will require not just borrowing models and corrections from elsewhere but adapting and testing them in ways that account for the islands' unique conditions.

II. Conclusion: Toward Next-Hour $PM_{2.5}$ Forecasting in Hawaiʻi

Air quality science has advanced dramatically in the past two decades, producing a deeper understanding of particulate matter, sharper tools for monitoring, and increasingly sophisticated forecasting models. Yet the literature also shows its limits. Many of the most widely cited studies are rooted in continental settings, where pollution stems primarily from industry and traffic. For Hawaiʻi, where volcanic emissions, marine aerosols, and trade wind dynamics define the atmosphere, the transferability of these findings is uncertain. This absence is not just an academic curiosity; it carries real implications for communities that live daily with vog and rely on timely information to make decisions about work, school, and health.

Calibration work with PurpleAir sensors has proven that low-cost devices can extend monitoring networks, but whether regression-based or deep learning corrections hold under Hawaiʻi's conditions is unknown. Meteorological research confirms the central role of boundary layer dynamics, wind, and humidity in shaping air quality, but few models explicitly capture how these variables interact within a small-island regime. Volcanic studies establish clear links between SO_2 emissions, sulfate aerosol formation, and human health outcomes, yet forecasting frameworks rarely integrate volcanic activity as a structured input. Advances in machine learning and hybrid modeling point toward more accurate predictions, but most remain prototypes that have not been applied in places with Hawaiʻi's distinctive atmospheric mix.

Taken together, these observations reveal three critical gaps that frame the present study. First, Hawaiʻi has been consistently excluded from calibration and forecasting research, leaving uncertainty about whether established correction factors or forecast models can be trusted in this setting. Second, much of the forecasting literature is oriented toward daily or multi-day horizons, while the timescale most relevant for residents (the next hour) remains underexplored. Third, volcanic emissions are typically treated as discrete hazards rather than integrated into continuous forecasting frameworks, despite their persistent influence on the islands' air quality.

The project proposed here directly addresses these gaps by testing whether simple, interpretable models can deliver skillful next-hour forecasts when supplied with three types of input: recent $PM_{2.5}$ history, basic meteorological variables, and an indicator of volcanic activity. Anchoring the analysis on AQS regulatory monitors ensures methodological rigor, while testing transferability to PurpleAir expands the potential reach of forecasts to neighborhoods where regulatory monitors are absent.

The contribution of this work is both local and global. For Hawaiʻi, it will provide a forecasting framework that is tuned to the realities of island meteorology and volcanic emissions, offering residents information that can shape everyday choices and protect vulnerable populations. For the broader field, it represents a test case of how forecasting methods perform when moved from their original continental settings into a maritime volcanic climate. That transferability question lies at the heart of applied environmental science: do models built in one place hold in another, or do they need to be reimagined to reflect local dynamics?

In this sense, the project does more than fill a gap in Hawaiʻi; it offers a broader challenge to the field of air quality science. If short-lead, interpretable models can work in one of the world's most complex atmospheric environments, they can likely be adapted elsewhere. If they cannot, then the field must confront the limits of its current approaches and invest in frameworks that account for regional diversity. Either way, the results from this study will not only inform local air quality management but also extend theoretical insight into how forecasting systems function across different environmental regimes.

List of References

Allabakash, S., & Lim, Y.-K. (2021). Climatology of planetary boundary layer height-controlling meteorological parameters over the Korean Peninsula. *Atmospheric Research, 259*, 105663. https://doi.org/10.1016/j.atmosres.2021.105663

Barkjohn, J., Bergin, M., Norris, G., Schauer, J., Kaimal, R., & Zhang, Y. (2021). Development and application of a United States-wide correction for $PM_{2.5}$ data collected with the PurpleAir sensor. *Atmospheric Measurement Techniques, 14*(6), 4617–4637. https://doi.org/10.5194/amt-14-4617-2021

Brook, R. D., Newby, D. E., Rajagopalan, S., Bard, R. L., Adar, S. D., & Wellenius, G. A. (2018). Volcanic smog and cardiometabolic health: Hawaiian hypertension? *Circulation, 138*(6), 666–668. https://doi.org/10.1161/CIRCULATIONAHA.118.035019

Businger, S., Huff, R., Horton, K. A., Sutton, A. J., Elias, T., & Thompson, M. (2015). Observing and forecasting vog dispersion from Kīlauea volcano, Hawaiʻi. *Bulletin of the American Meteorological Society, 96*(10), 1667–1686. https://doi.org/10.1175/BAMS-D-14-00150.1

Feng, X., Liang, H., & Zhang, J. (2024). A hybrid model for enhanced forecasting of $PM_{2.5}$ spatiotemporal concentrations with high resolution and accuracy. *Environmental Modelling & Software, 174*, 106300. https://doi.org/10.1016/j.envsoft.2024.106300

Ghahremanloo, M., Fountoukis, C., & Mantas, E. (2023). Deep learning calibration model for PurpleAir $PM_{2.5}$ measurements: Comprehensive investigation of the PurpleAir network. *Atmospheric Pollution Research, 14*(2), 101–115. https://doi.org/10.1016/j.apr.2022.101115

Gutierrez-Avila, I., Riojas-Rodriguez, H., & Grutter, M. (2021). Prediction of daily mean and one-hour maximum $PM_{2.5}$ concentrations and applications in central Mexico using satellite-based machine-learning models. *Atmospheric Environment, 252*, 118307. https://doi.org/10.1016/j.atmosenv.2021.118307

Kristiani, E., Tjahyadi, R., & Santoso, L. (2022). Short-term prediction of $PM_{2.5}$ using LSTM deep learning methods. *Applied Sciences, 12*(15), 7425. https://doi.org/10.3390/app12157425

Mathieu-Campbell, M., Heimburger, A., & Hagan, D. (2023). Calibration of PurpleAir low-cost particulate matter sensors: Model development for air quality under high relative humidity conditions. *Atmospheric Measurement Techniques, 16*(3), 621–635. https://doi.org/10.5194/amt-16-621-2023

Özupak, O., Mutlu, M., & Yılmaz, B. (2023). Air quality forecasting using machine learning: Comparative analysis and ensemble strategies for enhanced prediction. *Environmental Science and Pollution Research, 30*, 45678–45692. https://doi.org/10.1007/s11356-023-27123-2

Petrić, T., Ivković, B., & Djurić, I. (2021). Ensemble machine learning, deep learning, and time series forecasting: Improving prediction accuracy for hourly concentrations of $PM_{2.5}$ in America. *Atmospheric Pollution Research, 12*(9), 101157. https://doi.org/10.1016/j.apr.2021.101157

PurpleAir. (2022). *Particulate matter ($PM_{2.5}$) sensor loan program QAPP (AASB-QAPP-004-R1.2)*. PurpleAir, in partnership with EPA. https://www.epa.gov/air-research/pm25-sensor-loan-program

Saminathan, K., & Malathy, R. (2022). Ensemble-based classification approach for $PM_{2.5}$ concentration forecasting using meteorological data. *Environmental Monitoring and Assessment, 194*, 344. https://doi.org/10.1007/s10661-022-09936-1

Sayahi, T., Butterfield, Z., & Kelly, K. E. (2023). Long-term field evaluation of PurpleAir sensors for $PM_{2.5}$ monitoring. *Atmospheric Environment, 300*, 119748. https://doi.org/10.1016/j.atmosenv.2023.119748

Shi, X., Guo, J., & Li, J. (2021). Persistence baselines in air quality forecasting: Limits and applications. *Atmospheric Research, 249*, 105274. https://doi.org/10.1016/j.atmosres.2020.105274

Tam, E., Miike, R., Labrenz, S., Sutton, A., Elias, T., Davis, J., ... & Katz, A. (2016). Volcanic air pollution over the Island of Hawai‘i: Emissions, dispersal, composition, and association with respiratory symptoms and lung function. *Environmental International, 92–93*, 505–511. https://doi.org/10.1016/j.envint.2016.04.018

Tang, Y., Carmichael, G. R., & Thongboonchoo, N. (2018). Air quality impacts of the 2018 Mt. Kilauea volcano eruption in Hawai‘i: A regional chemical transport model study with satellite-constrained emissions. *Atmospheric Chemistry and Physics, 18*(22), 16155–16175. https://doi.org/10.5194/acp-18-16155-2018

Tofte, K., Giambelluca, T., & Businger, S. (2021). Large-scale weather patterns favorable for volcanic smog occurrences on O‘ahu, Hawai‘i. *Journal of Applied Meteorology and Climatology, 60*(3), 311–325. https://doi.org/10.1175/JAMC-D-20-0117.1

Tursumbayeva, A., Kenessov, B., & Baray, S. (2022). Planetary boundary layer and its relationship with $PM_{2.5}$ concentrations in Almaty, Kazakhstan. *Atmospheric Environment, 267*, 118756. https://doi.org/10.1016/j.atmosenv.2021.118756

U.S. Environmental Protection Agency. (2024). *2024 air monitoring network plan.* https://www.epa.gov/air-quality-management-processes/2024-air-monitoring-network-plan

U.S. Environmental Protection Agency. (2025). *5-year network assessment 06-30-final.* https://www.epa.gov/air-quality-management-processes/2025-network-assessment

U.S. Environmental Protection Agency. (2024). *Final reconsideration of the National Ambient Air Quality Standards for particulate matter.* https://www.epa.gov/pm-pollution/final-reconsideration-naaqs

U.S. Environmental Protection Agency. (2024). *Final updates to the Air Quality Index (AQI) for particulate matter: Fact sheet and common questions.* https://www.epa.gov/aqs/aqi-updates-2024

U.S. Geological Survey. (2017). *Volcanic air pollution hazards in Hawai'i.* U.S. Geological Survey Fact Sheet 2017–3017. https://doi.org/10.3133/fs20173017

Wang, J., Sun, Y., & Chen, J. (2019). Relationship analysis of $PM_{2.5}$ and boundary layer height using an aerosol and turbulence detection lidar. *Atmospheric Environment, 208*, 25–35. https://doi.org/10.1016/j.atmosenv.2019.04.006

Wang, L., Li, X., & Zhou, Y. (2023). Impact analysis of meteorological variables on $PM_{2.5}$ pollution in the most polluted cities in China. *Atmospheric Pollution Research, 14*(3), 101289. https://doi.org/10.1016/j.apr.2023.10128987

Wu, H., Zhao, Y., & Xu, M. (2022). Time-series data-driven $PM_{2.5}$ forecasting: From theoretical framework to empirical analysis. *Environmental Research Letters, 17*(12), 124014. https://doi.org/10.1088/1748-9326/ac91d9

Link to the References

Link to Google Drive Directory