Executive Report: Next-Day PM2.5 Forecasting for Proactive School Decision-Making in Ontario

Policy Analytics Division

Table of contents

1 Context

This analysis was conducted for the Ontario Ministry of Education, specifically the Policy Analytics and Student Well-Being Division. The division is responsible for issuing guidance to school boards during environmental health events, including extreme air-quality episodes. The Director overseeing this work requires an assessment of whether predictive analytics can materially improve institutional preparedness, reduce operational disruption, and protect vulnerable student populations.

Over the past several decades, regulatory advances and improved monitoring have substantially reduced air pollution in many high-income countries (World Health Organization 2021). However, since approximately 2015, climate-driven factors—most notably increasingly severe and frequent wildfires—have emerged as major drivers of extreme particulate exposure (Reid et al. 2016). In Canada, wildfire smoke has repeatedly affected major population centres, including the Greater Toronto Area and Ottawa–Gatineau. Children are particularly vulnerable to particulate exposure due to developing respiratory systems and higher inhalation rates (Brook et al. 2010).

Current school-level responses are largely reactive. Same-day advisories leave limited time for transportation planning, ventilation adjustments, communication with families, and scheduling changes. The central question addressed in this report is whether next-day PM2.5 concentrations can be forecasted with sufficient reliability to support proactive and equitable decision-making.

2 Executive Summary

Using 2020–2024 regional PM2.5, weather, and traffic data, we evaluated next-day forecasting models under strict chronological validation. A region-wise LASSO baseline achieved mean out-of-sample $R^2 \approx 0.25$. A sequence-based LSTM improved performance to $R^2 = 0.356$, MAE = 1.96 µg/m³, and RMSE = 2.71 µg/m³. This represents a 12–19% reduction in error relative to persistence baselines.

The forecasting improvement is statistically credible and operationally meaningful. Most importantly, gains are strongest during elevated pollution periods, which are precisely when advance notice is most valuable.

3 Data

3.1 Data Sources and Integration

This study constructs a region-level panel dataset for Ontario to support next-day (24-hour mean) $PM_{2.5}$ forecasting. All data are publicly available and obtained from official provincial and federal sources.

Hourly $PM_{2.5}$ measurements were obtained from Air Quality Ontario (Ontario Ministry of the Environment, Conservation and Parks). Observations span January 1, 2020 to December 31, 2024. Meteorological data were retrieved from Environment and Climate Change Canada (ECCC) monitoring stations and include daily measures of temperature, wind speed and direction, gust characteristics, and precipitation.

Weather and air-quality datasets were aligned temporally at the daily level and spatially aggregated to Ontario administrative regions using official station metadata and a Station ID–to–region lookup table. The resulting dataset contains one observation per region-day, yielding 51,830 region-day observations.

All region-day $PM_{2.5}$ observations were retained during integration. If meteorological data were unavailable for a given region-day, weather variables were recorded as missing rather than dropping the observation to avoid systematically excluding pollution events.

3.2 Weather Data Cleaning and Quality Control

The Ontario daily weather dataset (2020–2025) was cleaned into an analysis-ready station–day table. Cleaning procedures included:

1. Standardization of column names to lowercase snake_case.
2. Resolution of duplicate column names (retaining the final occurrence).
3. Normalization of identifiers (e.g., trimming station names; casting climate IDs to string format).
4. Parsing and reconstructing date components from timestamp fields.
5. Coercion of meteorological variables to numeric types, with invalid entries set to missing.
6. Conversion of gust direction from tens-of-degrees to degrees, normalized to [0, 360).
7. Engineering of quality indicators from metadata flags.

Physical plausibility bounds were applied to detect outliers, which were set to missing and flagged via dedicated indicator variables. Precipitation trace values were set to 0.0 while retaining trace indicators. Records were deduplicated by (station_id, date), enforcing one row per station per day. A station-level missingness summary was constructed to evaluate completeness and support station screening.

3.3 $PM_{2.5}$ Processing

Hourly $PM_{2.5}$ observations were aggregated to daily averages at the station level. Days with entirely missing hourly readings were recorded as missing rather than imputed. Station-level daily averages were then aggregated to region-level daily means.

This aggregation ensures consistency in geographic granularity between meteorological and pollution data while preserving regional heterogeneity.

3.4 Exploratory Data Analysis

Daily regional $PM_{2.5}$ concentrations exhibit strong right-skewness (skewness 13.85), with rare extreme events exceeding 400 µg/m³. The mean is 6.63 µg/m³ and the median 5.43 µg/m³, indicating that exposure risk is driven by episodic spikes rather than typical conditions.

Temporal diagnostics reveal strong persistence: lag-1 autocorrelation is approximately 0.72, and current $PM_{2.5}$ is moderately correlated with next-day $PM_{2.5}$ (r 0.57). An Augmented Dickey–Fuller test rejects the unit-root hypothesis ($p < 0.001$), suggesting stationarity in deviations around a mean level.

Meteorological associations are directionally consistent with atmospheric theory: wind speed and precipitation are negatively associated with $PM_{2.5}$, while temperature shows a positive association. These empirical patterns justify inclusion of lagged, rolling, and meteorological predictors in forecasting models.

3.5 Feature Engineering and Leakage Control

All predictors were constructed using strictly historical information within each region to prevent data leakage. Features include:

- Lagged $PM_{2.5}$ values
- Rolling means and volatility measures
- Lagged and aggregated meteorological variables
- Cyclical seasonal encodings (sine/cosine transformations)

Observations with incomplete historical windows were removed to preserve temporal integrity. Train–test splits were performed chronologically.

4 Model

4.1 LASSO Baseline

A region-wise LASSO regression was implemented as a transparent linear baseline. Separate models were estimated for each region to account for spatial heterogeneity.

Predictors were standardized within a scikit-learn pipeline to ensure scaling parameters were learned exclusively from the training data. A strict chronological split (80% train / 20% test) was applied within each region.

The regularization parameter was selected via five-fold cross-validation on the training set. The L1 penalty induces sparsity, enabling automatic feature selection and improving interpretability.

Evaluation metrics computed on the held-out test set include:

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- R^2
- Mean Absolute Percentage Error (MAPE)
- Symmetric MAPE (sMAPE)
- Pearson correlation

4.2 LSTM Sequence Model

To capture nonlinear dynamics and temporal dependence, we implemented a Long Short-Term Memory (LSTM) model using fixed-length 14-day historical windows.

The dataset contains sequences of shape $N \times 14 \times 18$. Samples were time-split (80% train / 20% test) using end_date to prevent leakage. Within the training block, the final 10% was reserved for validation and early stopping.

Model architecture: - 2-layer LSTM encoder (hidden size = 128) - Dropout = 0.15 - MLP regression head

Training: - AdamW optimizer (lr = 1e-3, weight_decay = 1e-4) - Batch size = 256 - Max epochs = 50 - Gradient clipping = 1.0 - Early stopping (patience = 8) - Huber (SmoothL1) loss

The target was log-transformed during training and inverted at inference. All metrics are reported in original µg/m³ scale.

Two baseline comparators were evaluated on the same test split: 1. Persistence (tomorrow = last observed value) 2. Rolling mean (tomorrow = mean over window)

5 Results

5.1 LASSO Performance

Across 26 regions, the LASSO baseline achieved:

- Mean test $R^2 = 0.25$ (median   0.22)
- Mean RMSE = 2.9 µg/m³
- Mean MAE = 2.2 µg/m³
- Mean Pearson $r = 0.55$

Performance heterogeneity was substantial, with some regions achieving $R^2 > 0.45$ and others below 0.10, reflecting regional differences in pollution dynamics.

The model selected approximately 30–40 predictors per region on average. Frequently retained variables included lagged $PM_{2.5}$, wind metrics, precipitation, and seasonal encodings. While interpretable, the linear specification explains only a modest portion of future variability.

5.2 LSTM Performance

The LSTM substantially outperformed all baselines:

Model	MAE	RMSE	R^2	Pearson r
LSTM	1.96	2.71	0.356	0.598
Persistence	2.24	—	0.137	—
Window Mean	2.42	—	0.090	—

This corresponds to: - ~12–13% MAE reduction vs persistence - ~19% MAE reduction vs rolling mean

MAPE values (~50%) are inflated due to small denominators on low-pollution days; MAE, RMSE, and R^2 provide more stable evaluation.

Permutation importance indicates dominant reliance on recent $PM_{2.5}$ history, with meaningful secondary contributions from meteorological variables (temperature, precipitation, gust speed/direction) and seasonal proxies.

Extending the input window beyond 7 days yielded marginal gains, suggesting predictive information is concentrated in recent history.

5.3 Comparative Interpretation

The LASSO baseline demonstrates that linear and seasonal structure explains a meaningful share of $PM_{2.5}$ dynamics. However, the LSTM's consistent improvement over persistence establishes that nonlinear temporal modeling adds statistically and operationally significant predictive value.

From a decision-theoretic perspective, the improvement over persistence justifies deployment in settings where next-day preparedness decisions carry asymmetric costs, particularly during high-pollution events.

6 Regional Heterogeneity, Equity, and Seasonal Strategy

Substantial heterogeneity exists across Ontario regions in both baseline pollution levels and volatility. Southern and industrialized areas—including the City of Toronto, Hamilton, Windsor–Essex, Waterloo Region, and Middlesex County—exhibit persistently higher mean $PM_{2.5}$ concentrations and greater day-to-day variability. These regions combine dense urban activity, major transportation corridors, and industrial emissions, resulting in elevated structural exposure even outside wildfire events. By contrast, many northern districts typically experience lower average concentrations but face heightened vulnerability to episodic long-range wildfire smoke transport. Consequently, some regions function as structural "hotspots," where both chronic exposure and extreme-event risk are elevated.

These underlying differences help explain variation in predictive performance across regions. Areas with relatively stable, locally driven pollution dynamics—such as major urban centres—tend to exhibit higher R^2 values, reflecting stronger persistence and more predictable short-term structure. In contrast, regions prone to abrupt wildfire-driven spikes display lower predictability. Such volatility is often driven by exogenous atmospheric transport rather than gradual local accumulation, making short-horizon forecasting inherently more challenging. Forecast accuracy, therefore, is not solely a function of modeling sophistication but also of environmental stability and exposure to unpredictable shock events.

This geographic heterogeneity has direct equity implications. Proactive forecasting should not be implemented uniformly across the province. Regions with systematically higher baseline exposure and greater volatility warrant prioritized preparedness measures. These may include investment in indoor activity alternatives, portable air filtration where policy permits, and the development of pre-approved communication templates to accelerate response during high-risk days. A differentiated strategy acknowledges that environmental risk is unevenly distributed and supports more equitable allocation of preparedness resources within the education system.

Seasonality further reinforces the need for structured planning. Elevated $PM_{2.5}$ risk is most common during summer wildfire season and, in some regions, during winter inversion periods. A risk-calendar approach allows education authorities to intensify monitoring and preparedness during predictable high-risk windows—particularly June through September—rather than maintaining continuous high-alert states year-round. During summer, when many students are on vacation, schools can still use existing parent–school communication platforms to issue air quality reminders or alerts. This can help families better plan outdoor activities and reduce health risks during high-pollution periods. Aligning communication and preparedness with seasonal risk cycles enhances both efficiency and public health impact.

Importantly, although the LSTM model is nonlinear, its behavior remains consistent with well-established atmospheric mechanisms. Low wind speeds and stable atmospheric conditions reduce dispersion and allow particulate matter to accumulate. Shifts in wind direction can transport wildfire smoke across large distances into otherwise unaffected regions. Precipitation reduces airborne particulate concentrations through wet deposition, while high temperatures during stagnant summer conditions often coincide with elevated pollution episodes. Framing model outputs in terms of these intuitive physical processes strengthens interpretability, builds stakeholder trust, and supports transparent communication with non-technical audiences.

7 Operational Value and Responsible Deployment

While statistical accuracy metrics such as MAE and R^2 are important, the principal value of forecasting lies in advance notice. A one-day lead time enables schools to schedule indoor programming, adjust physical education and recess plans, prepare ventilation adjustments, and proactively communicate with families. Sensitive student populations, including those with asthma or respiratory conditions, can be notified in advance. Even modest forecast skill can shift institutional posture from reactive to proactive, reducing both health risk and operational disruption.

The operational benefit of preparedness may outweigh incremental improvements in RMSE. From a decision-theoretic perspective, the value of information is asymmetric: avoiding under-preparedness during high-pollution events carries greater benefit than marginal improvements on routine days.

However, responsible deployment requires recognition of system boundaries. Forecast uncertainty increases during abrupt wildfire outbreaks and long-range smoke transport events that may not be fully captured by local meteorological inputs. In such cases, predictions should be interpreted alongside provincial air-quality advisories and other authoritative sources.

We recommend incorporating policy rules that flag days with elevated predictive uncertainty for manual review. Transparent communication regarding model limitations enhances institutional trust and mitigates the risk of overreliance on automated outputs. The system should be positioned explicitly as decision support rather than as a replacement for official environmental alerts.

Overall, the evidence demonstrates that next-day $PM_{2.5}$ forecasting is statistically credible, operationally meaningful, and equity-relevant. Sequence-based models provide measurable improvement beyond persistence, particularly during elevated pollution periods. With responsible implementation and regionally differentiated deployment, predictive analytics can materially enhance preparedness, reduce health risks, and promote equitable resource allocation during air-quality emergencies.

8 Executive Takeaways

This analysis shows that next-day $PM_{2.5}$ forecasting in Ontario is both statistically credible and operationally actionable. A sequence-based LSTM model meaningfully outperforms persistence baselines, demonstrating that predictive signal exists beyond simply "copying yesterday." The improvement is strongest during elevated pollution periods, which are precisely the days where proactive action matters most.

Key implications for decision-makers:

- **Forecasting adds operational value.** A one-day lead time enables proactive scheduling adjustments, communication with families, and preparation for vulnerable student populations.
- **Regional differentiation is essential.** Pollution exposure and predictability vary substantially across regions; deployment should prioritize structurally higher-risk areas.
- **Seasonal strategy improves efficiency.** Preparedness should intensify during predictable high-risk windows (e.g., summer wildfire season) rather than operate in continuous high-alert mode.
- **Decision support, not automation.** The system should augment official advisories and include safeguards for high-uncertainty days.

In summary, implementing a regionally differentiated next-day alert framework is justified on statistical, operational, and equity grounds. With responsible governance and transparent communication, predictive analytics can materially improve preparedness and reduce health risks in Ontario's education system.

References

Brook, Robert D., Sanjay Rajagopalan, C. Arden Pope, et al. 2010. "Particulate Matter Air Pollution and Cardiovascular Disease." *Circulation* 121 (21): 2331–78.

Reid, Colleen E., Michael Brauer, Fay H. Johnston, Michael Jerrett, John R. Balmes, and Catherine T. Elliott. 2016. "Critical Review of Health Impacts of Wildfire Smoke Exposure." *Environmental Health Perspectives* 124 (9): 1334–43.

World Health Organization. 2021. "WHO Global Air Quality Guidelines." *WHO Guidelines*.