BIBLIOGRAPHIC INFORMATION SYSTEM

JOURNAL FULL TITLE: Journal of Biomedical Research & Environmental Sciences
ABBREVIATION (NLM): J Biomed Res Environ Sci **ISSN:** 2766-2276 **WEBSITE:** https://www.jelsciences.com

SCOPE & COVERAGE

- **Sections Covered:** 34 specialized sections spanning 143 topics across Medicine, Biology, Environmental Sciences, and General Science
- Ensures broad interdisciplinary visibility for high-impact research.

PUBLICATION FEATURES

- **Review Process:** Double-blind peer review ensuring transparency and quality
- **Time to Publication:** Rapid 21-day review-to-publication cycle
- **Frequency:** Published monthly
- **Plagiarism Screening:** All submissions checked with iThenticate

INDEXING & RECOGNITION

- **Indexed in:** Google Scholar, IndexCopernicus (**ICV 2022: 88.03**)
- **DOI:** Registered with CrossRef (**10.37871**) for long-term discoverability
- **Visibility:** Articles accessible worldwide across universities, research institutions, and libraries

OPEN ACCESS POLICY

- Fully Open Access journal under Creative Commons Attribution 4.0 License (CC BY 4.0)
- Free, unrestricted access to all articles globally

GLOBAL ENGAGEMENT

- **Research Reach:** Welcomes contributions worldwide
- **Managing Entity:** SciRes Literature LLC, USA
- **Language of Publication:** English

SUBMISSION DETAILS

- Manuscripts in Word (.doc/.docx) format accepted

SUBMISSION OPTIONS

- **Online:** https://www.jelsciences.com/submit-your-paper.php
- **Email:** support@jelsciences.com, support@jbresonline.com

HOME ABOUT ARCHIVE SUBMIT MANUSCRIPT APC

Vision: The Journal of Biomedical Research & Environmental Sciences (JBRES) is dedicated to advancing science and technology by providing a global platform for innovation, knowledge exchange, and collaboration. Our vision is to empower researchers and scientists worldwide, offering equal opportunities to share ideas, expand careers, and contribute to discoveries that shape a healthier, sustainable future for humanity.

JOURNAL OF BIOMEDICAL RESEARCH & ENVIRONMENTAL SCIENCES
ISSN: 2766-2276
SUBJECT AREA(S): EPIDEMIOLOGY | PUBLIC HEALTH
DOI: 10.37871

MINI REVIEW

Uncertainty-Aware Machine Learning for Ambient Air-Pollution Exposure Surfaces in Biomedical Research: From Data Fusion to Neuroepidemiology-Ready Inference

Betekhtin AA*

ITMO University, Lomonosova St., 9, 191002, Saint Petersburg, Russia

Abstract

Ambient air pollution remains a major, preventable driver of cardio metabolic and neurological disease burden. For biomedical studies, the central methodological bottleneck is not only prediction of pollutant concentrations, but trustworthy exposure assessment: leakage-safe validation, Uncertainty Quantification (UQ), transportable models in low-monitor regions, and transparent propagation of exposure uncertainty into health-effect estimates. This mini-review synthesizes recent advances in global and regional PM2.5 mapping, spatiotemporal deep learning, virtual monitoring stations, and gap-filling, and links these developments to the rapidly expanding evidence on dementia risk. We provide a practical checklist and worked calculations that translate modern Machine Learning (ML) exposure products into epidemiology-ready inputs.

Key Points (What Biomedical Reviewers Usually Look for)

- **Exposure surfaces:** ML models must be evaluated with *spatial* and *temporal* cross-validation that matches the target use (e.g., out-of-region prediction), not only random splits [1,2].
- **Uncertainty:** point predictions are insufficient; credible intervals (or full predictive distributions) are needed to propagate exposure error into health-effect inference [3].
- **Transportability:** hybrid "physics + ML" approaches and

***Corresponding author(s)**

Betekhtin AA, ITMO University, Lomonosova St., 9, 191002, Saint Petersburg, Russia

Email: aabetekhtin@itmo.ru

DOI: 10.37871/jbres2245

Submitted: 12 December 2025

Accepted: 29 December 2025

Published: 30 December 2025

Keywords

- Air pollution
- Exposure modelling
- Machine learning
- Uncertainty quantification
- Spatiotemporal deep learning
- $PM_{2.5}$
- NO_2
- Dementia
- Epidemiology

VOLUME: 6 ISSUE: 12 - DECEMBER, 2025

Scan Me

How to cite this article: Betekhtin AA. Uncertainty-Aware Machine Learning for Ambient Air-Pollution Exposure Surfaces in Biomedical Research: From Data Fusion to Neuroepidemiology-Ready Inference. J Biomed Res Environ Sci. 2025 Dec 30; 6(12): 1996-2001. doi: 10.37871/jbres2245, Article ID: JBRES2245, Available at: https://www.jelsciences.com/articles/jbres2245.pdf

geophysical priors reduce degradation far from monitors.

- **Open data:** harmonized monitoring streams (e.g., OpenAQ) and standardized metadata improve reproducibility, but versioning and API changes must be documented [4].
- **Biomedical relevance:** recent systematic reviews and large cohorts support associations between long-term pollution exposure and incident dementia, motivating higher-resolution and better-validated exposure models [5-9].

Introduction (Why "ML for air quality" is Now a Biomedical Methods Topic)

The 2021 WHO Global Air Quality Guidelines substantially tightened recommended levels for key pollutants, including $PM_{2.5}$ (Annual mean 5μg/m³; 24-hour 15μg/m³) [10]. In Europe, updated indicators continue to report a large burden attributable to $PM_{2.5}$ exposures [2]. Regulatory tightening (e.g., the EU recast Ambient Air Quality Directive) and new accountability mechanisms (Including legal avenues for affected citizens) increase demand for transparent, uncertainty-aware evidence [11-15].

For biomedical research, the key deliverable is an exposure surface: a spatial–temporal field $x(\mathbf{s},t)$ that can be linked to participants by location history. Modern surfaces are typically produced by data fusion (Monitors + satellite AOD + chemical transport models + meteorology + land use) and increasingly by spatiotemporal deep learning [16-18] However, an exposure model that minimizes mean squared error can still be *unsafe* for epidemiology if it leaks information across space/time, fails in low-monitor regions, or provides no UQ.

Core Definitions (Terms Used Consistently in This Paper)

- **Exposure surface** $x(\mathbf{s},t)$: Estimated pollutant concentration at location **s** and time t, aligned to the health-study time scale (Daily, monthly, annual).
- **Data fusion**: Combining multiple information sources (Monitors, satellites, CTMs, land-use predictors) to estimate $x(\mathbf{s},t)$ [18,19].
- **Spatial cross-validation**: Validation that withholds entire regions (or monitors) to test transportability; contrasts with random splits that can overestimate performance [1,2].
- **Uncertainty quantification (UQ)**: Reporting predictive uncertainty (e.g., standard deviation $\sigma(\mathbf{s},t)$ or predictive intervals) and propagating it into downstream analyses [3].

What the Last Wave of Global $PM_{2.5}$ ML Mapping Changed (2019-2025)

Three trends dominate recent high-impact exposure modelling:

Global, long-term $PM_{2.5}$ fields with consistent methodology

High-resolution, long-term global $PM_{2.5}$ products now combine satellites, models, and monitors with statistical/ML layers, enabling decade-scale exposure assessment [16-18]. These surfaces are attractive for cohort studies because they offer wide coverage and consistent back-casting.

"Physics + ML" to improve low-monitor transportability

Purely data-driven models often degrade

far from monitors. Incorporating geophysical *a priori* estimates into deep learning explicitly targets this failure mode [1]. The implication for biomedical studies is straightforward: improved out-of-sample performance reduces differential exposure misclassification between urban (Monitor-rich) and rural (Monitor-sparse) participants.

Epidemiology-facing UQ and reproducibility

Methodological work increasingly emphasizes uncertainty-aware fusion and explicit validation protocols [3]. In parallel, open monitoring infrastructures facilitate reproducible pipelines, but only if API versions, licensing, and provenance are recorded [4,20].

Practical checklist for an epidemiology-ready ML exposure model

Table 1 summarizes failure modes that frequently trigger reviewer pushback.

Worked Examples / Calculations (With Sanity Checks)

Example 1: Exceedance probability using a predictive distribution

Suppose an ML surface provides, for a given day and location, a predictive mean μ and standard deviation σ for daily $PM_{2.5}$. To estimate the probability of exceeding the WHO 24-hour guideline $g = 15\mu g/m^3$, a simple (Often used) approximation is a normal predictive distribution:

$$P(\text{exceed}), \approx 1 - \Phi\left(\frac{g-\mu}{\sigma}\right), \quad (1)$$

where Φ is the standard normal CDF.

Numerical example (units and sanity check). Let $\mu = 12\mu g/m^3$ and $\sigma = 4\mu g/m^3$. Then

$$z = \frac{g-\mu}{\sigma} = \frac{15-12}{4} = 0.75, \qquad P$$

(Exceed) $\approx 1 - \Phi(0.75) \approx 1 - 0.773 = 0.227$.

Sanity check: since $\mu < g$, exceedance probability should be < 0.5; 22.7% is plausible.

Example 2: Attenuation of a health-effect estimate by classical exposure error

Let the (Unobserved) true long-term exposure be X^* and the estimated exposure be $X = X^* + \varepsilon$ with independent noise ε. In classical measurement error, regression coefficients are attenuated approximately by

$$\lambda = \frac{\mathrm{Var}(X^*)}{\mathrm{Var}(X^*) + \mathrm{Var}(\varepsilon)}, \quad (2)$$

Thus, a "true" association β^* may be observed as $\beta \approx \lambda\beta^*$. This is a central motivation for UQ and transportability-focused modelling.

Numerical example. Assume between-person

Table 1: Epidemiology-ready checklist for ML exposure surfaces.

Item	What to report / do
Target time scale	Define *t* (daily / monthly / annual) and justify for disease latency (e.g., dementia: multi-year means) [5,6]
Spatial CV	Report region-holdout / monitor-holdout performance (Not only random CV) [1,2]
Uncertainty	Provide predictive intervals or distributions; show calibration (Coverage) [3]
Data provenance	Document monitoring sources and versions (e.g., OpenAQ v3; retired v1/v2 endpoints) [4]
Missingness	Describe gap-filling strategy for monitors/time series if used [25]
Non-stationarity	Address trend/drift (Policy changes, emissions shifts) in training/validation [18]
Leakage controls	Ensure no future data inform past predictions; avoid spatial "bleed" from nearby monitors in random splits [2]

long-term exposure variability SD(X^*) = 6μg/m³, so Var(X^*) = 36. If the exposure model has RMSE ≈ 3μg/m³, a rough proxy is Var(ε) ≈ 9. Then

$$\lambda = \frac{36}{36+9} = 0.80$$

Sanity check: better models (Smaller RMSE) increase λ toward 1, reducing attenuation.

Example 3: Monte Carlo propagation of exposure uncertainty into a Cox model

When an exposure surface provides (μ_i,σ_i) for participant *i*, a simple uncertainty-propagation workflow is:

1. For *m* = 1,...,*M* draws, sample $X_i^{(m)} \sim \mathcal{N}(\mu_i, \sigma_i^2)$ (or use the model's predictive distribution).
2. Fit the health model (e.g., Cox) to each draw to obtain $\hat{\beta}^{(m)}$.
3. Report the distribution of $\hat{\beta}^{(m)}$ (mean, CI), separating *statistical* uncertainty from *exposure* uncertainty.

Why Dementia is a Compelling "biomedical endpoint" for ML Exposure Methods

The evidence base linking long-term ambient pollution to incident dementia has expanded rapidly in recent years. A 2025 systematic review and meta-analysis synthesized the growing observational literature [21], complementing earlier broad syntheses. Large cohort studies report associations between long-term $PM_{2.5}$/NO_2 exposure and dementia/Alzheimer's disease incidence. Mechanistically adjacent neurodegenerative outcomes are also being investigated; for example, a 2025 *Science* study reported links between long-term $PM_{2.5}$ exposures and Lewy body dementia.

For such endpoints, the methodological requirement is stronger than for short-latency outcomes: multi-year averaging, sensitivity analyses to mobility, and robust out-of-region exposure prediction become essential. Hence, "physics + ML" transportability gains and UQ are not cosmetic features; they directly affect bias and interpretability.

Emerging Methods That Reviewers Now Expect You to Cite

Beyond global mapping, biomedical submissions increasingly cite:

- **Forecasting architectures** that couple decomposition + graph learning + sequence models (Useful for short-term health endpoints and operational warnings).
- **Virtual monitoring stations** that estimate concentrations in unmonitored locations using ML (Relevant when residential geocoding is fine-grained).
- **Gap-filling benchmarks** for incomplete monitoring time series (Important if you build local fusion models from raw monitors).
- **Map recovery / sparse sensing** concepts that formalize reconstruction from limited sensors.
- **Policy context** that motivates thresholds and public-health interpretation (WHO guidelines; EU Directive 2024/2881) [22-31].

Conclusion

Machine learning has shifted ambient air-pollution exposure assessment from coarse averages to high-resolution, global and regional surfaces. For biomedical research, the next bar is *trust*: spatially honest validation, calibrated uncertainty, and transparent propagation of exposure error into health models. These requirements align with regulatory tightening and a rapidly growing neuroepidemiology literature on dementia risk. A pragmatic path for

submissions in ML-focused biomedical journals is to present exposure modelling as an *inference pipeline* rather than a pure prediction task: data provenance (e.g., OpenAQ), transportability (Physics + ML), UQ, and sensitivity analyses that match the disease time scale.

Data Availability Statement

This mini-review used publicly accessible documentation and published literature. No new human subject data were collected.

References

1. Agbehadji IE, Obagbuwa IC. Systematic review of machine learning and deep learning techniques for spatiotemporal air quality prediction. Atmosphere. 2024;15:1352. doi: 10.3390/atmos15111352.
2. Shen S, Li C, van Donkelaar A, Jacobs N, Wang C, Martin RV. Enhancing Global Estimation of Fine Particulate Matter Concentrations by Including Geophysical a Priori Information in Deep Learning. ACS EST Air. 2024 Mar 27;1(5):332-345. doi: 10.1021/acsestair.3c00054. PMID: 38751607; PMCID: PMC11092969.
3. Malings C, Knowland KE, Pavlovic N, Coughlin JG, King D, Keller C, Cohn S, Martin RV. Air quality estimation and forecasting via data fusion with uncertainty quantification: Theoretical framework and preliminary results. JGR: Machine Learning and Computation. 2024. doi: 10.1029/2024JH000183.
4. OpenAQ docs. About the API. OpenAQ API. 2025.
5. Best Rogowski CB, Bredell C, Shi Y, Tien-Smith A, Szybka M, Fung KW, Hong L, Phillips V, Jovanovic Andersen Z, Sharp SJ, Woodcock J, Brayne C, Navaratnam A, Khreis H. Long-term air pollution exposure and incident dementia: a systematic review and meta-analysis. Lancet Planet Health. 2025 Jul;9(7):101266. doi: 10.1016/S2542-5196(25)00118-4. Epub 2025 Jul 24. PMID: 40716448.
6. Shi L, Steenland K, Li H, Liu P, Zhang Y, Lyles RH, Requia WJ, Ilango SD, Chang HH, Wingo T, Weber RJ, Schwartz J. A national cohort study (2000-2018) of long-term air pollution exposure and incident dementia in older adults in the United States. Nat Commun. 2021 Nov 19;12(1):6754. doi: 10.1038/s41467-021-27049-2. PMID: 34799599; PMCID: PMC8604909.
7. Andersen ZJ, Lim YH, Zhang J, STuffier S, Cole-Hunter T, Bergmann M, Loft S, Mortensen LH, Chen J, Stafoggia M, de Hoogh K, Katsouyanni K, Vienneau D, Rodopoulou S, Samoli E, Bauwelinck M, Klompmaker JO, Atkinson R, Janssen NAH, Oftedal B, So R. Long-term exposure to air pollution and risk of dementia among older individuals of a Danish nationwide administrative cohort. Environment International. 2025. doi: 10.1016/j.envint.2025.109607.
8. Mortamais M, Gutierrez LA, de Hoogh K, Chen J, Vienneau D, Carrière I, Letellier N, Helmer C, Gabelle A, Mura T, Sunyer J, Benmarhnia T, Jacquemin B, Berr C. Long-term exposure to ambient air pollution and risk of dementia: Results of the prospective Three-City Study. Environ Int. 2021 Mar;148:106376. doi: 10.1016/j.envint.2020.106376. Epub 2021 Jan 20. PMID: 33484961.
9. Zhang X, Liu H, Wu X, Jia L, Gadhave K, Wang L, Zhang K, Li H, Chen R, Kumbhar R, Wang N, Terrillion CE, Kang BG, Bai B, Park M, Denna MCF, Zhang S, Zheng W, Ye D, Rong X, Yang L, Niu L, Ko HS, Peng W, Jin L, Ying M, Rosenthal LS, Nauen DW, Pantelyat A, Kaur M, Irene K, Shi L, Feleke R, García-Ruiz S, Ryten M, Dawson VL, Dominici F, Weber RJ, Zhang X, Liu P, Dawson TM, Han S, Mao X. Lewy body dementia promotion by air pollutants. Science. 2025 Sep 4;389(6764):eadu4132. doi: 10.1126/science.adu4132. Epub 2025 Sep 4. PMID: 40906862; PMCID: PMC12459341.
10. WHO global air quality guidelines: Particulate matter (PM2.5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide. Geneva: WHO. 2021.
11. European parliament. Air pollution: Parliament adopts revised law to improve air quality. 2025.
12. European commission. New pollution rules come into effect for cleaner air by 2030. 2025.
13. Directive (EU) 2024/2881 of the European parliament and of the council of 23 October 2024 on ambient air quality and cleaner air for Europe (Recast). Official Journal of the European Union. Available via EUR-Lex. 2024.
14. Reuters. EU strikes deal to strengthen air quality standards. 2025.
15. Reuters. EU Parliament adopts new rules to improve air quality by 2030. 2025.
16. Yu W, Ye T, Zhang Y, Xu R, Lei Y, Chen Z, Yang Z, Zhang Y, Song J, Yue X, Li S, Guo Y. Global estimates of daily ambient fine particulate matter concentrations and unequal spatiotemporal distribution of population exposure: a machine learning modelling study. Lancet Planet Health. 2023 Mar;7(3):e209-e218. doi: 10.1016/S2542-5196(23)00008-6. PMID: 36889862.

17. van Donkelaar A, Hammer MS, Bindle L, Brauer M, Brook JR, Garay MJ, Hsu NC, Kalashnikova OV, Kahn RA, Lee C, Levy RC, Lyapustin A, Sayer AM, Martin RV. Monthly Global Estimates of Fine Particulate Matter and Their Uncertainty. Environ Sci Technol. 2021 Nov 16;55(22):15287-15300. doi: 10.1021/acs.est.1c05309. Epub 2021 Nov 1. Erratum in: Environ Sci Technol. 2024 Mar 5;58(9):4463-4464. doi: 10.1021/acs.est.4c01477. PMID: 34724610.

18. Hammer MS, van Donkelaar A, Li C, Lyapustin A, Sayer AM, Hsu NC, Levy RC, Garay MJ, Kalashnikova OV, Kahn RA, Brauer M, Apte JS, Henze DK, Zhang L, Zhang Q, Ford B, Pierce JR, Martin RV. Global Estimates and Long-Term Trends of Fine Particulate Matter Concentrations (1998-2018). Environ Sci Technol. 2020 Jul 7;54(13):7879-7890. doi: 10.1021/acs.est.0c01764. Epub 2020 Jun 17. PMID: 32491847.

19. van Donkelaar A, Martin RV, Li C, Burnett RT. Regional Estimates of Chemical Composition of Fine Particulate Matter Using a Combined Geoscience-Statistical Method with Information from Satellites, Models, and Monitors. Environ Sci Technol. 2019 Mar 5;53(5):2595-2611. doi: 10.1021/acs.est.8b06392. Epub 2019 Feb 12. PMID: 30698001.

20. OpenAQ Docs. Measurements resource (Purpose, aggregation, access patterns). 2025.

21. Premature deaths due to exposure to fine particulate matter in Europe. European Environment Agency (EEA). 2025.

22. Song J, Fan H, Gao M, Xu Y, Ran M, Liu X, Guo Y. Toward high-performance map-recovery of air pollution data from sparse monitoring networks. ACS ES&T Engineering. 2022. doi: 10.1021/acsestengg.2c00248.

23. Wang X, Zhang S, Chen Y, He L, Ren Y, Zhang Z, Li J, Zhang S. Air quality forecasting using a spatiotemporal hybrid deep learning model based on VMD-GAT-BiLSTM. Sci Rep. 2024 Aug 1;14(1):17841. doi: 10.1038/s41598-024-68874-x. PMID: 39090177; PMCID: PMC11294351.

24. Makhdoomi A, Sarkhosh M, Ziaei S. PM2.5 concentration prediction using machine learning algorithms: an approach to virtual monitoring stations. Sci Rep. 2025 Mar 8;15(1):8076. doi: 10.1038/s41598-025-92019-3. PMID: 40057563; PMCID: PMC11890590.

25. Safarov R, Shomanova Z, Nossenko Y, Kopishev E, Bexeitova Z, Kamatov R. Filling gaps in PM2.5 time series: A broad evaluation from statistical to advanced neural network models. PLoS One. 2025 Aug 14;20(8):e0330211. doi: 10.1371/journal.pone.0330211. PMID: 40811692; PMCID: PMC12352854.

26. Indicator EN.ATM.PM25.MC.M3: Population weighted exposure to ambient PM2.5 (Definition and methodology). World Bank (WDI Metadata Glossary). 2025.

27. Wilker EH, Osman M, Weisskopf MG. Ambient air pollution and clinical dementia: systematic review and meta-analysis. BMJ. 2023 Apr 5;381:e071620. doi: 10.1136/bmj-2022-071620. PMID: 37019461; PMCID: PMC10498344.

28. Kulick ER, Wellenius GA, Boehme AK, Joyce NR, Schupf N, Kaufman JD, Mayeux R, Sacco RL, Manly JJ, Elkind MSV. Long-term exposure to air pollution and trajectories of cognitive decline among older adults. Neurology. 2020 Apr 28;94(17):e1782-e1792. doi: 10.1212/WNL.0000000000009314. Epub 2020 Apr 8. PMID: 32269113; PMCID: PMC7274848.

29. Bernacki J, Scherer R. A Comprehensive Review of Data-Driven Techniques for Air Pollution Concentration Forecasting. Sensors (Basel). 2025 Oct 1;25(19):6044. doi: 10.3390/s25196044. PMID: 41094865; PMCID: PMC12526560.

30. Rajesh M, Babu RG, Moorthy U, Easwaramoorthy SV. Machine learningdriven framework for realtime air quality assessment and predictive environmental health risk mapping. Sci Rep. 2025 Aug 6;15(1):28801. doi: 10.1038/s41598-025-14214-6. PMID: 40770019; PMCID: PMC12328577.

31. Im U, Ye Z, Schuhen N, Chowdhury S, Christensen JH, Geels C, Hänninen R, Hodnebrog O, Marelle L, Sofiev, Brandt MJ, Aunan K. Europe will struggle to meet the new WHO air quality guidelines under plausible emission scenarios. npj Clean Air. 2025.