

Spatially Resolved Wildfire Smoke AQI Forecasting with Physics-Informed Transformers

Emilie Beck
Andrei Bachinin
emilie.beck@colostate.edu
andrei.bachinin@colostate.edu
Colorado State University
Fort Collins, USA

Parker Jones
Brenna Wolf
parker.jones@colostate.edu
brenna.wolf@colostate.edu
Colorado State University
Fort Collins, USA

Sangmi Lee Pallickara
Shrideep Pallickara
sangmi.pallickara@colostate.edu
shrideep.pallickara@colostate.edu
Colorado State University
Fort Collins, USA

Abstract

Wildfire smoke produces spatially extensive air-quality degradation, but ground-truth AQI measurements are available only at sparse monitoring locations. This paper introduces PROMETHEUS, a physics-guided framework for spatially resolved wildfire-smoke AQI forecasting. PROMETHEUS harmonizes MODIS and VIIRS fire detections, wildfire perimeters, GridMet meteorology, 30-meter fuel representations derived from NLCD and Anderson fuel models, and EPA air-quality observations into a common fire-centered learning representation. To address the mismatch between sparse sensor supervision and dense AQI prediction, PROMETHEUS augments observational loss with science-guided spatial regularizers and, for multi-day forecasting, an advection-diffusion residual that constrains predicted AQI fields to evolve consistently with smoke transport. Experiments over 5,095 fires show that PROMETHEUS improves pointwise accuracy and produces more physically coherent smoke-impact fields than CNN, CNN+LSTM, and Prithvi baselines.

CCS Concepts

• **Computing methodologies** → **Neural networks**; Supervised learning; • **Information systems** → **Spatial-temporal systems**.

Keywords

physics-informed deep learning, ViT, multimodal data fusion

ACM Reference Format:

Emilie Beck, Andrei Bachinin, Parker Jones, Brenna Wolf, Sangmi Lee Pallickara, and Shrideep Pallickara. 2026. Spatially Resolved Wildfire Smoke AQI Forecasting with Physics-Informed Transformers. In *The 34th ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '26)*, November 03–06, 2026, Riverside, CA, USA. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3841645.3844206>

1 Introduction

Wildfires in the United States have grown in frequency, intensity, and duration. Development has pushed into the wildland-urban interface, exposing more people. Smoke reaches far beyond the fire itself, carrying fine particulate matter, carbon monoxide, nitrogen and sulfur oxides, and ozone precursors; exposure causes respiratory and cardiovascular harm. The U.S. Environmental Protection

Agency (EPA) summarizes these pollutants in the Air Quality Index (AQI), which grades conditions from Good (0–50) to Hazardous (301–500), so predicting wildfire-driven AQI degradation is an urgent public-health priority.

Existing approaches each solve part of the problem: station-level machine-learning models return scalar time series rather than continuous fields, chemical transport models demand detailed emissions inventories and heavy computation, and wildfire risk models characterize fire behavior rather than downwind AQI. None delivers a dense, continuously updated map of smoke-driven AQI degradation that evolves with the fire.

The obstacle is a structural mismatch between a dense prediction target and sparse supervision. The EPA operates more than 13,000 AQI sensors, but they cluster near populated areas while wildfires originate in remote terrain, so on a given fire-day ground truth exists at fewer than 1% of the pixels a model must predict. Trained only at observed locations, a model interpolates well near sensors but produces incoherent fields elsewhere: fragmented plumes and implausible gradients. Fixing this requires encoding what is already known about how smoke moves, so PROMETHEUS uses physics not as a mechanistic simulator but as an inductive bias that regularizes learning where observations are missing.

Research Questions. Our overarching question is: *How can we model the spatiotemporal dynamics of air quality in regions affected by forest fires in a way that accounts for the evolving footprint of the fire itself and the influencing environmental factors?* Within this, we explore:

RQ-1: *How can we construct a more comprehensive dataset that captures the spatial and temporal evolution of wildfire spread?*

RQ-2: *How can we incorporate topographical and meteorological variables (vegetation, elevation, wind, humidity, and temperature) to improve our understanding of fire behavior and pollutant dispersion?*

RQ-3: *How can we model fine-grained variations in air quality, both across space and over time?*

2 Related Work

Machine-learning models based on LSTMs, ensembles, and deep networks forecast AQI accurately at fixed stations, but emit scalar time series and ignore wildfire-specific drivers such as burn-area progression, fuel load, and wind transport [4]. Chemical transport models take the opposite approach: CMAQ solves the governing equations for advection, diffusion, and chemical transformation, but needs detailed emissions inventories and heavy computation [1]. Physics-informed neural networks sit between the



This work is licensed under a Creative Commons Attribution 4.0 International License. *SIGSPATIAL '26, Riverside, CA, USA*

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2950-8/2026/11

<https://doi.org/10.1145/3841645.3844206>

two by embedding partial differential equations in the training loss, letting a model respect known physical laws where observations are sparse [6]; the same paradigm has been applied to geospatial fields, coupling multimodal sensing with process-based modeling to map soil moisture at high resolution from sparse ground truth [5]. Vision Transformers supply the architectural ingredient, since self-attention over patches captures long-range dependencies that convolutions recover poorly [2]. PROMETHEUS bridges these lines: it learns a dense, fire-centered AQI field with transformer encoders, disciplines it with science- and physics-guided losses, and is benchmarked against CNN, CNN+LSTM, and Prithvi-EO baselines [3].

3 Methodology

3.1 Data Construction & Harmonization [RQ-1]

We combine three archives: NASA FIRMS detections of active-combustion fire pixels from MODIS and VIIRS; MTBS burned-area polygons; and the Spatial Wildfire Occurrence archive of discovery dates, sizes, and center points. The latter two give us *where* and *when* each fire burned, letting us attribute detections to a specific event-day. Daily detections are fragmented, so for each fire-day we approximate the active perimeter with the convex hull of that day’s fire pixels; in sequence the hulls trace the fire’s progression. A hull can overestimate area, but it guarantees that all detected activity falls inside the perimeter, the right bias for an emission source.

Around each hull we draw a 160-km buffer defining the sampling domain, the range over which AQI readings typically attenuate to near-background levels. Within it we align AQI readings from the EPA/AirNow network of more than 13,000 sensors, averaged per pollutant over ozone, PM2.5, PM10, CO, SO₂, and NO₂ — the sparse supervision target; ten GridMet fields spanning burning index, temperature, precipitation, humidity, radiation, and wind; and a 30-meter fuel grid mapping NLCD land cover onto Anderson’s 13 fire-behavior fuel models. All are burned onto one co-registered lattice, so every pixel encodes fire extent, emission potential, atmospheric forcing, and any observed pollutant response. We restrict the study to the contiguous United States, where the Anderson models are calibrated; after removing fire-days without AQI coverage we retain 5,095 fires spanning 20,691 fire-days.

3.2 Physics-Guided Architectures for Sparsely Supervised AQI Field Prediction [RQ-2/3]

We address two tasks. *Nowcasting* predicts a dense AQI field for a single day; *forecasting* predicts a sequence of fields over a future horizon. Each has a data-driven and a physics-guided variant, giving *Prometheus-Nowcast*, *Prometheus-Nowcast-Sci*, *Prometheus-Forecast*, and *Prometheus-Forecast-Sci*. The difficulty is identical in all four: covariates are spatially dense but supervision is not, so a model trained only at sensor pixels can reach a reasonable pointwise error while producing implausible fields elsewhere.

3.2.1 Shared Spatial Backbone. Each fire-day is a 32-channel raster tensor: 10 GridMet bands, a fuel band, a binary fire-mask band marking the hull, 18 pollutant-summary bands (mean, maximum, and minimum of each pollutant), and two metadata bands for log-normalized fire size and day of year. We split the tensor into non-overlapping 8×8 patches, project each into a 256-dimensional

embedding with a strided convolution, add learned positional embeddings, and process the sequence with transformer blocks of self-attention and feed-forward layers. Attention is essential here, since a pixel’s AQI depends jointly on the fire perimeter, wind, terrain, fuel, and distant sensors. The patch features are then reshaped to a feature grid, bilinearly upsampled to the input resolution, and refined by a convolutional decoder into a dense field $\hat{y} \in \mathbb{R}^{H \times W}$.

3.2.2 Nowcasting with Science-Guided Losses. *Prometheus-Nowcast* trains on a masked Huber loss, $\mathcal{L}_{\text{Huber}} = \frac{1}{\sum_i m_i} \sum_{i:m_i=1} \mathcal{L}_\delta(\hat{y}_i - y_i)$ with mask $m_i \in \{0, 1\}$, evaluated only where ground truth exists, since treating unmeasured pixels as zeros would impose false supervision. The Huber form, quadratic below $\delta = 1.0$ and linear above, limits the pull of extreme AQI values. But it constrains the model only at monitors.

Prometheus-Nowcast-Sci adds three soft regularizers over the whole predicted field. A full transport model on a single snapshot would be underdetermined, so these express softer preferences. Let $\langle f \rangle_w = \sum_i w_i f_i / \sum_i w_i$ be the proximity-weighted mean, with $w_i \in [0, 1]$ decaying exponentially with distance to the nearest station, so constraints act most strongly where observations give local support; write $[z]_+ = \max(0, z)$. Smoothness $\mathcal{L}_s$ penalizes excess local curvature through a five-point discrete Laplacian. Fire-distance attenuation $\mathcal{L}_g$ encodes that concentration decreases away from active fire pixels, with d the distance to the nearest fire pixel and ∇_a the central-difference gradient along axis $a \in \{x, y\}$; it penalizes only violations, so displaced plumes are never forced into radial symmetry. Range regularization $\mathcal{L}_b$ holds predictions within normalized bounds $[y_{\min}, y_{\max}] = [-2, 4]$ standard deviations without clipping high-AQI values:

$$\begin{aligned} \mathcal{L}_s &= \langle (\nabla^2 \hat{y})^2 \rangle_w, \\ \mathcal{L}_g &= \left\langle \sum_a [\nabla_a \hat{y} \cdot \text{sign}(\nabla_a d)]_+^2 \right\rangle_w, \\ \mathcal{L}_b &= \langle [y_{\min} - \hat{y}]_+^2 + [\hat{y} - y_{\max}]_+^2 \rangle_w. \end{aligned} \quad (1)$$

The combined objective is $\mathcal{L}_{\text{total}} = \lambda_d \mathcal{L}_{\text{Huber}} + \alpha(\lambda_s \mathcal{L}_s + \lambda_g \mathcal{L}_g + \lambda_b \mathcal{L}_b)$, with a curriculum factor $\alpha \in [0, 1]$ that rises linearly from zero so the model first learns the empirical relationship among fire, fuel, weather, and AQI. These terms do not replace data; they extend its reach into pixels the Huber loss never sees.

3.2.3 Forecasting and Physics Fine-Tuning. *Prometheus-Forecast* slides a window of length $T_{\text{hist}} + T_{\text{fore}}$ over each fire, using $T_{\text{hist}} = 7$ history days to predict $T_{\text{fore}} = 7$ future maps. A shallower spatial encoder embeds each history day, and we reshape the result so every spatial location becomes its own time series, preventing the map from collapsing into one global sequence, which would encourage spatially uniform forecasts. A temporal encoder-decoder transformer ($D_t = 512$, 6 layers, 8 heads) models that evolution using one learnable query per forecast day, causal self-attention, and cross-attention to the encoded history; without the causal mask the queries attend bidirectionally and collapse toward similar representations. Training uses a horizon-weighted objective: the w_t -weighted mean of the per-day masked Huber losses, so days can be weighted by operational importance.

Prometheus-Forecast-Sci carries the principal physics-guided component. Once consecutive predicted fields exist, we can ask not only

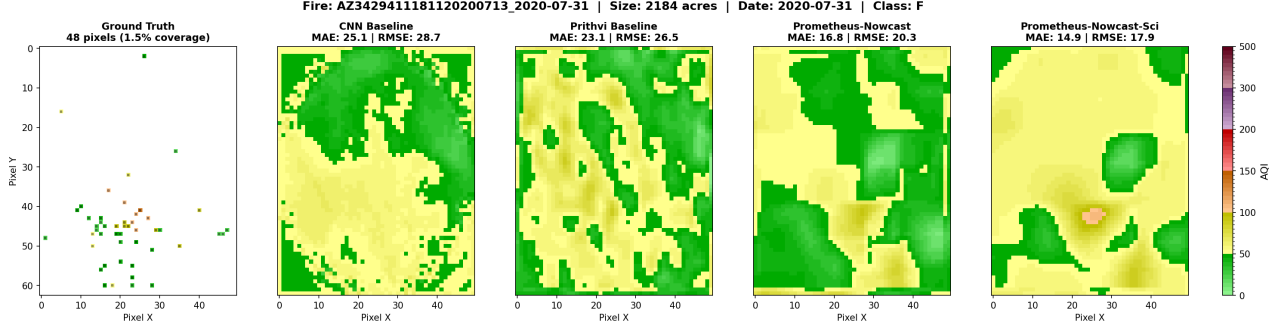


Figure 1: Single-day AQI predictions on a held-out fire-day, left to right: ground truth station readings; the *CNN baseline*, fitting sparse sensor pixels but returning blocky, piecewise output; the *Prithvi baseline* (frozen Prithvi-EO-1.0 encoder, trainable adapter and decoder), competitive pointwise but limited in dynamic range; *Prometheus-Nowcast* (ViT), globally consistent through self-attention across patches; and *Prometheus-Nowcast-Sci* (ViT + Science), the most physically coherent dispersion pattern. Per-panel MAE/RMSE (AQI points) confirm that Prometheus also improves on both baselines at the measured pixels.

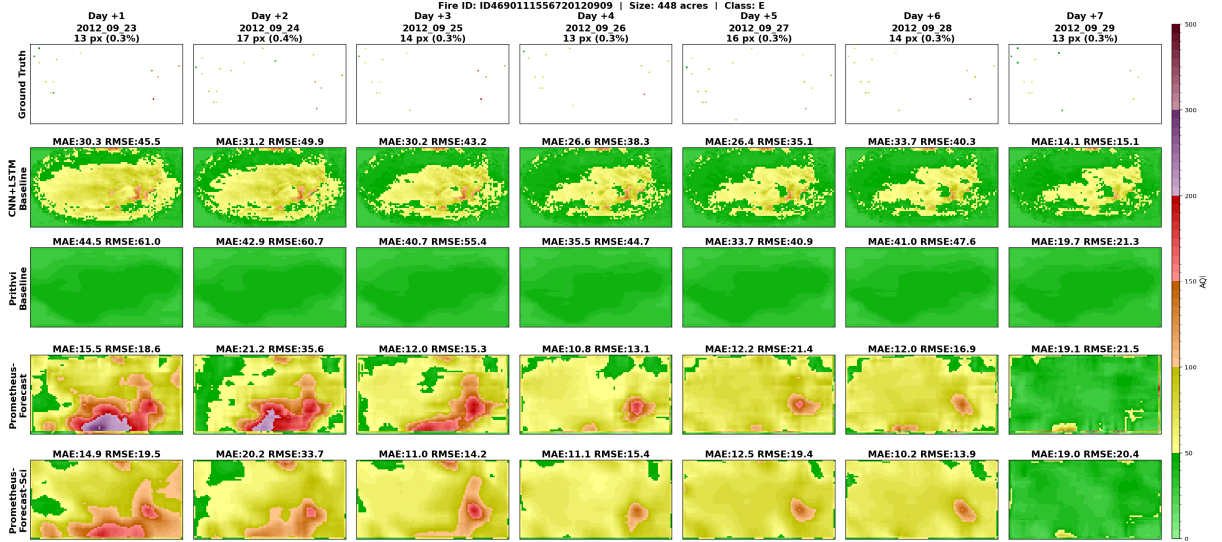


Figure 2: Seven-day AQI forecasts on a held-out fire event, by row: ground truth station readings; the *CNN+LSTM baseline*, fragmented because a convolutional encoder with a per-location recurrent head cannot model long-range spatial dependencies; the *Prithvi baseline* (frozen Prithvi-EO-1.0, trainable temporal head), which collapses to near-mean forecasts that fall entirely inside the EPA *Good* band (AQI 0–50) on this event, so the row renders uniformly green under the discrete AQI colormap even though its raw outputs vary slightly between pixels and days; *Prometheus-Forecast* (Temporal ViT), spatially coherent and evolving across the horizon; and *Prometheus-Forecast-Sci* (Physics Fine-Tuning), which refines those forecasts with advection–diffusion constraints, reducing artifacts.

whether each matches its sensors, but whether the transition between them is consistent with atmospheric transport. Let C denote pollutant concentration at position (x, y) and time t . Its evolution approximates the advection–diffusion balance

$$\frac{\partial C}{\partial t} + u \cdot \nabla C = D \nabla^2 C - kC + S, \quad (2)$$

where $u = (u_x, u_y)$ is the wind velocity vector, D a diffusion coefficient, k a first-order decay rate representing deposition and chemical removal, and S a source term injecting smoke at active

fire locations. Rearranging Equation 2 gives a residual for the transition between consecutive days t and $t + 1$,

$$R = \frac{C_{t+1} - C_t}{\Delta t} + u_x \frac{\partial C}{\partial x} + u_y \frac{\partial C}{\partial y} - D \nabla^2 C + kC - S, \quad (3)$$

which vanishes when the predicted sequence obeys the balance. Spatial derivatives come from C_t by central differences with a five-point Laplacian stencil, the temporal derivative from C_{t+1} with $\Delta t = 1$ day, $S = s \cdot F$ from the binary fire mask F , and $\mathcal{L}_{\text{pde}}$ is the mean of R^2 over interior pixels, excluding a two-pixel border where finite differences are unreliable. The parameters D , k , and

s are dimensionless in grid-cell units, scaled so the residual stays commensurate with the Huber term.

Two properties make this effective. The first is an asymmetry: $\mathcal{L}_{\text{Huber}}$ is sparse because it needs measurements, whereas $\mathcal{L}_{\text{pde}}$ is dense for forecast-to-forecast transitions, where both fields are model predictions. The physics term supplies training signal over exactly the domain the sensors miss, and its coefficient is kept small so it regularizes without dominating. The second is staging. Training with the residual from the start proved ineffective, because it competes with the model's initial effort to learn the relationship among fire, weather, fuel, and AQI. We train *Prometheus-Forecast* to convergence on the horizon-weighted Huber objective, then fine-tune with $\mathcal{L}_{\text{ft}} = \lambda_h \mathcal{L}_{\text{Huber}} + \lambda_p \mathcal{L}_{\text{pde}}$, so physics refines an already competent forecaster rather than competing with it.

4 Performance Benchmarks & Discussion

This evaluation is asymmetric: the models predict a dense field, but supervision exists only at sparse monitors. MAE and RMSE therefore quantify fidelity at observed pixels, while Figures 1 and 2 address whether the field shows *spatially coherent smoke structure* consistent with fire location, meteorology, and transport. Since a model can lower MAE by interpolating near stations while failing between them, we read both together. We train in PyTorch on an A100 with AdamW and early stopping, partitioning at the fire-event level so no wildfire appears in both training and test.

Single-Day AQI Nowcasting. The CNN baseline, a convolutional encoder–decoder over the same rasterized drivers, achieves MAE 14.96 and RMSE 19.47, and the Prithvi baseline, a trainable head on the frozen 100M-parameter Prithvi-EO-1.0 encoder [3], achieves MAE 14.45 and RMSE 18.79. *Prometheus-Nowcast* improves on both at MAE 11.96 and RMSE 15.94, and *Prometheus-Nowcast-Sci* attains MAE 12.06 and RMSE 16.06 — nearly the same pointwise accuracy, but the most physically coherent field. Figure 1 shows why the visual axis is needed: the CNN returns blocky regions unlike transported smoke and Prithvi misses the plume, while science-guided regularization improves field realism at minimal cost to pointwise error.

Multi-Day AQI Forecasting. The CNN+LSTM baseline, a per-day convolutional encoder with a per-pixel recurrent head, spans MAE 17.28–17.78 and RMSE 28.19–29.18 across the seven-day horizon; the Prithvi temporal baseline, reusing the frozen encoder under a trainable temporal transformer, spans MAE 18.02–18.41 and RMSE 29.88–31.07. *Prometheus-Forecast* achieves MAE 14.23–14.51 and RMSE 23.82–24.15, improving on both at every forecast day, and *Prometheus-Forecast-Sci*, fine-tuned for 29 epochs with a 2.5% physics-loss contribution, reaches MAE 13.97–14.17 and RMSE 23.63–23.90. The gains concentrate late in the horizon, where Day+6 and Day+7 MAE drop from 14.33 and 14.36 to 14.01 and 14.10, so the residual is most useful once the data-driven model accumulates artifacts. Figure 2 agrees: the CNN+LSTM row is fragmented, the Prithvi row falls almost entirely inside the EPA *Good* band, and the PROMETHEUS rows show coherent dispersion evolving over the horizon, indicating that the temporal transformer captures field evolution rather than extrapolating station-level readings.

5 Conclusions

PROMETHEUS treats wildfire-smoke AQI forecasting as spatial field prediction rather than station-level regression.

RQ-1: Daily convex-hull perimeters over MODIS and VIIRS fire pixels give a consistent, interpretable representation of wildfire evolution across 5,095 fires and 20,691 fire-days, making each fire-day a spatially explicit event whose geometry evolves.

RQ-2: A 30-meter fuel grid from NLCD and Anderson fuel models, with GridMet forcings, lets the model reason jointly about smoke generation, fuel availability, and dispersion.

RQ-3: Cast as dense field reconstruction under sparse supervision, the transformer models capture long-range dependencies among fire extent, fuel, meteorology, and AQI, outperforming CNN, CNN+LSTM, and Prithvi baselines; the physics-guided variants further constrain the unobserved field through science-guided regularizers for single days and an advection–diffusion residual during fine-tuning.

Both dimensions improve: attention captures the nonlocal dependencies the CNN baselines miss, while the physics-guided losses supply structure where supervision is absent. Physics-guided learning is valuable whenever supervision is sparse but the deployment target is dense.

Acknowledgments

This research was supported by the National Science Foundation (2312319, 2617041), the National Institute of Food Agriculture (2025-77039-45531, COL014021223), an NSF/NIFA Artificial Intelligence Institutes AI-LEAF Award [2023-03616], and a Clare Boothe Luce Professorship.

References

- [1] Daewon Byun and Kenneth L. Schere. 2006. Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Applied Mechanics Reviews* 59, 2 (2006), 51–77. doi:10.1115/1.2128636
- [2] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xi-aohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. 2021. An image is worth 16×16 words: Transformers for image recognition at scale. In *Proceedings of the 9th International Conference on Learning Representations (ICLR)*. arXiv:2010.11929.
- [3] Johannes Jakubik, Sujit Roy, C. E. Phillips, Paolo Fraccaro, Denys Godwin, Bianca Zadrozny, Daniela Szwarcman, Carlos Gomes, Gabby Nyirjesy, Blair Edwards, et al. 2023. Foundation models for generalist geospatial artificial intelligence. arXiv preprint arXiv:2310.18660.
- [4] R. Janarthanan, P. Partheeban, K. Somasundaram, and P. Navin Elamparithi. 2021. A deep learning approach for prediction of air quality index in a metropolitan city. *Sustainable Cities and Society* 67 (2021), 102720. doi:10.1016/j.scs.2021.102720
- [5] Paahuni Khandelwal, Andrei Bachinin, Tanjim Bin Faruk, Everett Lewark, Sangmi Lee Pallickara, and Shrideep Pallickara. 2026. DeepSoil: Physics-Guided Learning for High-Resolution Soil Moisture Mapping from Surface to Root-Zone via Multimodal Sensing and Process-Based Modeling. *ACM Trans. Spatial Algorithms Syst.* 12, 4, Article 22 (July 2026), 39 pages. doi:10.1145/3817120
- [6] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics* 378 (2019), 686–707.