

Innovative Methods for Multi-source Air Dust Pollution Monitoring using Remote Sensing: A Comparative Study of Chalkidiki (Greece) and Ihalainen (Finland)

Petra Kopecká Sedláčková¹

petra.sedlackova@geology.cz

Czech Geological Survey

Kostas Gounaris²

Eldorado Gold Greece, Hellas Gold Single member SA

Veronika Kopačková Strnadová¹

Czech Geological Survey

Article

Keywords: CAMS, MODIS, PM10, dust pollution, mining area, time series analysis

Posted Date: September 21st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10915354/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Innovative Methods for Multi-source Air Dust Pollution Monitoring using Remote Sensing: A Comparative Study of Chalkidiki (Greece) and Ihalainen (Finland)

Petra Kopecká Sedláčková^{1*}, Kostas Gounaris², Veronika Kopačková Strnadová¹

¹Czech Geological Survey, ²Eldorado Gold Greece, Hellas Gold Single member SA

Corresponding author*: Petra Kopecká Sedláčková, Czech Geological Survey, Klárov 131/3, 118 00 Praha 1, Czech Republic, email: petra.sedlackova@geology.cz

Abstract

Mining activities can cause episodic and persistently elevated particulate pollution, yet consistent monitoring is challenging where ground networks are limited. We quantified PM₁₀ conditions at two mining-impacted study areas—Chalkidiki (Greece) and Ihalainen/Lappeenranta (Finland)—from 2013 to 2024 by combining ground measurements with multi-source remote sensing -based atmospheric products (CAMS and MODIS) across the two regions, PM₁₀ concentrations were higher in Greece and showed strong variability between monitoring stations, indicating pronounced local influences on dust loading. In the intercomparison against in situ observations, CAMS PM₁₀ provided the most reliable representation of temporal changes, yielding a strong station-based correspondence (Spearman $r_s \approx 0.8$). Long-term evaluation further indicated a decreasing PM₁₀ trajectory at both locations over the study period, suggesting gradual improvements in regional air quality. To assess operational usefulness, we conducted forecasting experiments and found that model performance depends on timescale: Prophet captured long-term tendencies most effectively, while XGBoost outperformed for short-term variability. Overall, the findings demonstrate that integrating ground observations with CAMS and complementary satellite information enables scalable monitoring of mining-related dust impacts.

Keywords

CAMS, MODIS, PM₁₀, dust pollution, mining area, time series analysis

1 Introduction

The composition of the Earth's atmosphere is complex, consisting of gases and aerosols [1], that include both liquid and solid particles of different origins and sizes [2,3]. Natural sources have historically contributed most to the atmospheric dust pollution [1] referred as total suspended particulate (TSP) or particulate matter (PM) [4]. However, in recent decades, anthropogenic activities have become dominant in producing large quantities of particles that accumulate in the lower atmosphere [2].

Particles are usually classified into three categories according to their size: PM₁₀ (coarse particles <10 μm), PM_{2.5} (fine particles <2.5 μm) and PM_{0.1} (ultrafine particles <0.1 μm) [5]. Fine and ultrafine particles are more likely to be combustion products, whereas coarse particles usually include mechanically generated dust [6], such as that produced by mining activities. Elevated concentrations of PM are of concern because of their ability to penetrate the human respiratory tract. Short-term exposure can cause irritation and coughing, while long-term exposure is linked to higher incidence of asthma, cardiovascular disease, and premature mortality [7,8]. Beyond health impacts, deposited dust can impair vegetation by blocking photosynthesis and transpiration [8], and on a global scale, aerosols

play a role in radiative forcing and cloud formation, thereby influencing weather and climate [9]. For these reasons, monitoring atmospheric dust and providing early warnings are essential.

In order to protect public health, the European Union introduced Directive 2008/50/EC [10], which establishes limit values for PM₁₀ concentrations. These are a daily average of 50 µg/m³ that should not be exceeded more than 35 times per year and an annual mean of 40 µg/m³. However, regions with mining activities in particular often have higher concentrations of PM in the air due to dust emissions and resuspension, creating a need for continuous monitoring [8].

The conventional approach to monitoring dust pollution relies on in-situ stations [11]. These stations are typically operated by the local meteorological office or by another organization responsible for site-specific air quality, and they provide regular interval records of air sample composition. Although these data are accurate and precise, their main limitation is their restricted spatial coverage and representation, which prevents spatially continuous monitoring. Establishing dense monitoring networks is often costly and logistically challenging. [11,12].

Complementary optical remote sensing approaches offer broad spatial coverage through regular grid-based observations and often provide high temporal resolution. While these data do not directly measure ground-level pollutant concentrations, they record the scene's physical properties by measuring the radiation it emits or reflects at specific wavelengths. This creates parameters that are strongly related to the presence and variability of particulate matter [3,13–15].

At the same time, air quality assessments also rely on numerical modelling approaches that allow several data inputs to be evaluated and complex air pollution scenarios to be created. Based on spatial scales, local models are designed to resolve sharp concentration gradients caused by traffic or industrial facilities, but they strongly depend on detailed emission inventories and dense meteorological input. Regional models, such as Copernicus Atmosphere Monitoring Service (CAMS), have wider spatial coverage and require fewer input variables, but tend to smooth out local variability and underestimate local extremes. While these products cannot replace local monitoring, they can provide an essential framework to complement sparse monitoring networks, without incurring the cost of terrain analysis [16].

In this study, we tested and validated freely available satellite data products and services, demonstrating their suitability for monitoring the spatial and temporal dynamics of dust emissions from open-pit mining. More specifically, we evaluated MODIS-derived Aerosol Optical Depth (AOD) retrievals and combined aerosol products from the CAMS. MODIS AOD (Terra/Aqua) provides columnar aerosol optical properties at multiple wavelengths with near-daily global coverage and moderate spatial resolution. AOD is a dimensionless optical parameter that quantifies the column-integrated extinction of solar radiation by aerosol particles due to combined scattering and absorption processes [15]. Formally, AOD (τ) is defined as the vertical integral of the aerosol extinction coefficient and it varies with wavelength, aerosol microphysical properties, and vertical distribution. Because AOD represents columnar burden rather than near-surface concentrations, it is commonly used as a proxy for particulate matter (PM) only after applying empirical or physically based relationships that account for aerosol vertical profile, humidity-driven hygroscopic growth, and aerosol type. AOD retrievals are also constrained by cloud contamination, surface-type effects (bright surfaces reduce retrieval accuracy), and retrieval algorithm assumptions (e.g., aerosol model selection and surface BRDF). These limitations reduce the fidelity of AOD-based PM estimations in urban, vegetated,

or frequently cloudy environments [13]. Moreover, existing studies [9,17–19] suggest that these products are more likely to be suitable for monitoring large dust events in arid regions.

The CAMS products are derived from numerical atmospheric composition models that assimilate multi-source observations — including spaceborne remote-sensing retrievals, surface in-situ measurements, and meteorological fields — using variational/data-assimilation techniques to provide analysis and reanalysis estimates of aerosol mass concentration and optical properties [16]. CAMS combined products provide gridded estimates of aerosol mass concentrations (e.g., PM_{2.5}, PM₁₀). These products explicitly simulate aerosol transport, deposition, and chemical transformation, thereby offering information on aerosol vertical distribution and source attribution that is absent from column-only satellite retrievals [16]. While CAMS is valuable for regional to global assessments and for providing temporally continuous fields (including during cloudy periods), the spatial resolution and model parameterizations may limit its accuracy in heterogeneous or complex terrain and for localized sources such as mine sites.

Therefore, the central research question addressed in this study is: **To what extent can CAMS reanalysis/analysis products and MODIS aerosol optical depth (AOD) support local PM₁₀ monitoring in mining areas with contrasting climatic conditions, such as temperate Finland and Mediterranean Greece?**

Accordingly, the specific objectives of this study are to:

- **Compare CAMS-derived PM₁₀ estimates with in situ PM₁₀ observations** at two climatically contrasting mining sites: Chalkidiki, Greece, and Ihalainen, Finland.
- **Assess the agreement between satellite/model-based products and ground-based measurements** by comparing CAMS PM₁₀ estimates and MODIS AOD with in situ PM₁₀ observations across the two study regions.
- **Characterize long-term trends and seasonality** by analysing multi-year temporal patterns, seasonal cycles, and site-specific drivers related to climate, geography, and mining activity.
- **Evaluate forecasting methods suitable for operational surveillance** by testing statistical and machine-learning approaches for short- and longer-term PM₁₀ prediction in support of routine air-quality monitoring near mining sites.
- **Provide a practical and reproducible workflow for resource-constrained monitoring** by developing an open and transferable approach based on freely available remote-sensing and atmospheric-modelling datasets, particularly for areas where conventional monitoring networks are sparse or absent.

2 Study area, Data and Methods

2.1 Study area

This study examines two contrasting case studies—Chalkidiki (Greece) and Lappeenranta (South Karelia, Finland)—selected to evaluate the robustness and transferability of mine-site dust modelling approaches. They span contrasting climates, dust source profiles, and dispersion regimes, enabling broader model generalizability. Chalkidiki’s Mediterranean setting combines hot, dry summers, sea-breeze circulations, complex terrain, and frequent Saharan dust intrusions with polymetallic mining activities [20], creating challenging conditions for source attribution and background correction. Lappeenranta’s boreal context features limestone/lime industrial operations with predominantly carbonate fugitive dust, strong seasonality (winter stability and spring road-dust peaks), and robust

Finnish monitoring/permitting frameworks that support validation [21]. Together, these sites test dust models across arid vs cold-humid conditions, metal-bearing vs carbonate dust chemistries, and different regulatory data environments, yielding findings that are both scientifically rigorous and policy-relevant.

2.1.1 Greece

The greek site is in the north-east of Chalkidiki, in Northern Greece (Central Macedonia), hosts long-lived mining districts centered on polymetallic sulfides (Au-Ag-Pb-Zn-Cu), skarn/porphyry systems (notably Skouries), and carbonate-hosted deposits (Stratoni–Madem Lakkos) [22,23]. Modern operations have been led by Hellas Gold S.A. (a subsidiary of Eldorado Gold) at three main sites: Olympias, Skouries, and Stratoni. Mining activities, located in distances from 1 to 5 km from populated areas, span underground and open-pit operations, ore processing (crushing, grinding, flotation), waste rock handling, tailings management, and haulage—each contributing potential air emissions, especially particulate matter (PM₁₀/PM_{2.5}) and metals-bearing dust [24–26].

2.1.2 Finland

The second site is in south-eastern Finland, in Ihalainen, near the city of Lappeenranta (South Karelia). In this region, industrial mineral operations are conducted by Nordkalk, a leading Northern European producer of limestone-based products. In and around Lappeenranta, Nordkalk operates limestone quarries and lime production facilities that supply raw and processed materials to regional industries (pulp and paper, construction, agriculture, flue-gas cleaning, and chemical processes) [27]. It is expected that PM₁₀ from quarrying and mineral handling is largely coarse, carbonate-rich dust. Fugitive dust from blasting, crushing, screening, conveying, unpaved roads, and stockpiles can raise near-source PM₁₀, impacts decay rapidly with distance and are mitigated by BAT controls (water sprays, enclosures, paving, wheel washing) [28].

Both test sites are together with in-situ stations visualised in following Fig. 1.

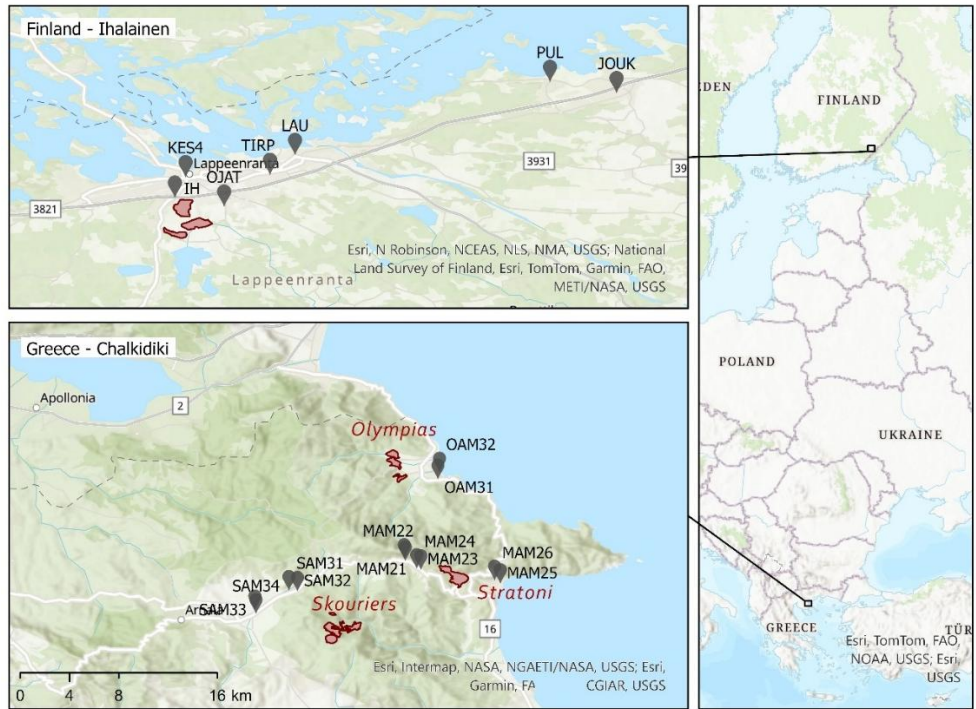


Fig. 1 Areas of interests - Quarries and in-situ monitoring locations at study areas in Finland and Greece

2.2 Data

In our study several types of data were analysed. First, in-situ PM₁₀ measurement datasets from both localities provided precise information in time and space, but their point-based nature limits their usefulness for spatial modelling. To complement this, remote sensing data and derived products were employed, offering much better spatial representation and coverage. On the other side, these EO-based datasets are constrained by the spatial resolution of the sensors and the processing chains used, which can limit their ability to capture fine spatial variability and continuous temporal patterns across the study period [15].

2.2.1 In-situ data

In-situ PM₁₀ data for Greece were obtained from the Hellas Gold environmental platform (<https://environmental.hellas-gold.com/>), which provides measurements of potential pollution from mining activities, including dust emissions. For this study, hourly PM₁₀ concentrations from residential areas were used, covering the period 06/2015–08/2024 at 12 monitoring locations.

In Finland, PM₁₀ data were sourced from the Finnish Meteorological Institute (FMI), which provides measurements of various air pollutants via its open data service (<https://en.ilmatieteenlaitos.fi/download-observations>). Here, hourly PM₁₀ observations from seven monitoring stations around Lappeenranta were used, spanning the period 01/2013–08/2024.

2.2.2 Data: Remote sensing products and services

2.2.2.1 MODIS Aerosol Optical Depth product data (AOD)

The MODIS Aerosol Optical Depth (AOD) dataset provides long-term continuity, twice-daily global sampling, and extensive validation through AERONET [14], making it suitable for large-scale climatological studies and for use as a proxy for surface particulate matter. Its applicability, however, can be constrained by retrieval gaps caused by clouds, snow, or sun glint, as well as by increased uncertainties over bright or heterogeneous land surfaces [15,18].

MODIS AOD products deliver global estimates of column-integrated aerosol loading, derived from radiances measured by the Moderate Resolution Imaging Spectroradiometer (MODIS) aboard NASA's Terra (morning overpass, since 1999) and Aqua (afternoon overpass, since 2002) satellites [18,29,30]. These products are generated using two complementary retrieval algorithms: the Dark Target (DT) approach, optimised for dark and vegetated surfaces, and the Deep Blue (DB) algorithm, designed for bright, highly reflective land such as deserts and urban areas [14]. Merged DT/DB products thus achieve near-global coverage and are available at multiple spatial resolutions, including 1–10 km (Level-2) and 1° (Level-3) [31].

In this study, we used Terra MODIS Level-2 aerosol products (Collection 6.1/6.2), which provide swath-based, geolocated AOD retrievals together with associated quality indicators at the time of satellite overpass. Specifically, AOD at 0.55 μm was extracted at three native resolutions: 10 km (MOD04_L2), 3 km (MOD04_3K), and 1 km (MCD19A2.061) (see Tab. 1).

	product	spatial resolution	temporal resolution	description
IN-SITU	Greece - PM ₁₀ in-situ monitoring data by Hellas Gold	points	hourly	12 monitoring stations, available from 6/2015
	Finland - PM ₁₀ in-situ monitoring data by FMI	points	hourly	7 monitoring stations, available from 1/2013
CAMS	European Air Quality Analysis data	0.1° x 0.1°	hourly	Ensemble numerical models of PM ₁₀ at surface level concentrations combining ground measurements and remote sensing data, available from 1/2021

	European Air Quality Reanalysis data	0.1° x 0.1°	hourly	Ensemble numerical models of PM ₁₀ at surface level concentrations combining ground measurements and remote sensing data, reanalysed for higher precision, available until 12/2020
MODIS	MOD04_L2 / MYD04_L2	10 km	4 times in day	Aerosol Optical Thickness at 0.55 µm for both Ocean (best) and Land (corrected) with best quality data (QA Confidence Flag = 3)
	MOD04_3K / MYD04_3K	3 km	4 times in day	Aerosol Optical Thickness at 0.55 µm for both Ocean (best) and Land (corrected) with best quality data (QA Confidence Flag = 3)
	MCD19A2.061	1 km	daily	Aerosol optical Depth at 0.55 µm, gridded Level 2 aerosol data derived from the Multi-Angle Implementation of Atmospheric Correction (MAIAC)

Tab. 1 Overview of data used in the study

2.2.2.2 Copernicus Atmosphere Monitoring Service (CAMS)

PM₁₀ concentrations at the surface were obtained from the Copernicus Atmosphere Monitoring Service (CAMS, ECMWF), which is part of the European Union's Copernicus Earth Observation programme. CAMS provides gridded PM₁₀ fields as global analyses and forecasts, as well as European regional ensemble products, generated with the ECMWF Integrated Forecasting System (IFS) chemistry–aerosol configuration and multi-source data assimilation of satellite and in-situ observations [32–35]. These products deliver dynamically consistent 3D aerosol and trace-gas fields suitable for background air-quality assessments, climatologies, and data fusion applications.

For this study, we used the CAMS Regional (Europe) ensemble reanalysis/forecast products at a horizontal resolution of approximately 0.1° (~10 km), provided on a regular latitude–longitude grid via the Atmosphere Data Store [36,37]. The underlying model outputs are available at hourly to three-hourly time steps in NetCDF and GRIB formats and include vertically resolved fields and surface-level mass concentrations of PM₁₀ (dry mass) in units of µg m⁻³. We extracted surface PM₁₀ ensemble mean fields, using the CAMS validated reanalysis dataset for the period 2013–2020 and the CAMS near-real-time analysis data for 2021–2024, with no additional bias correction applied beyond the standard CAMS quality control.

2.3 Data acquisition and preprocessing

Data acquisition and preprocessing were performed using Python (JupyterLab), R (RStudio), and ArcGIS Pro. In-situ PM₁₀ measurements were obtained directly from local monitoring platforms. CAMS surface PM₁₀ data were accessed via the Copernicus Climate Data Store (CDS) API, while MODIS AOD products were retrieved through NASA Earthdata services.

All datasets were processed in Python. For each ground-based station, the corresponding satellite and model grid values were extracted to construct point-based time series. All timestamps were converted to Coordinated Universal Time (UTC) to ensure temporal consistency across the different data sources. To harmonise the temporal resolution, the measurements in all datasets (CAMS, MODIS and in situ) were aggregated to daily medians. Missing values were retained and not interpolated in order to avoid introducing artificial temporal continuity, particularly in the AOD time series. For consistency with the temporal resolution of CAMS and the in-situ data, ground-based measurements were, where appropriate, further aggregated to “CAMS pixel-size” means (i.e., averages of all stations falling within a given model grid cell).

2.4 Methods

To assess dust (PM₁₀) pollution patterns and their temporal evolution at the two study sites, we applied a multi-step analytical framework combining exploratory statistics, multivariate analysis, correlation assessment, time series decomposition, predictive modelling, and spatial analysis. Visual methodology is provided in Fig. 2.

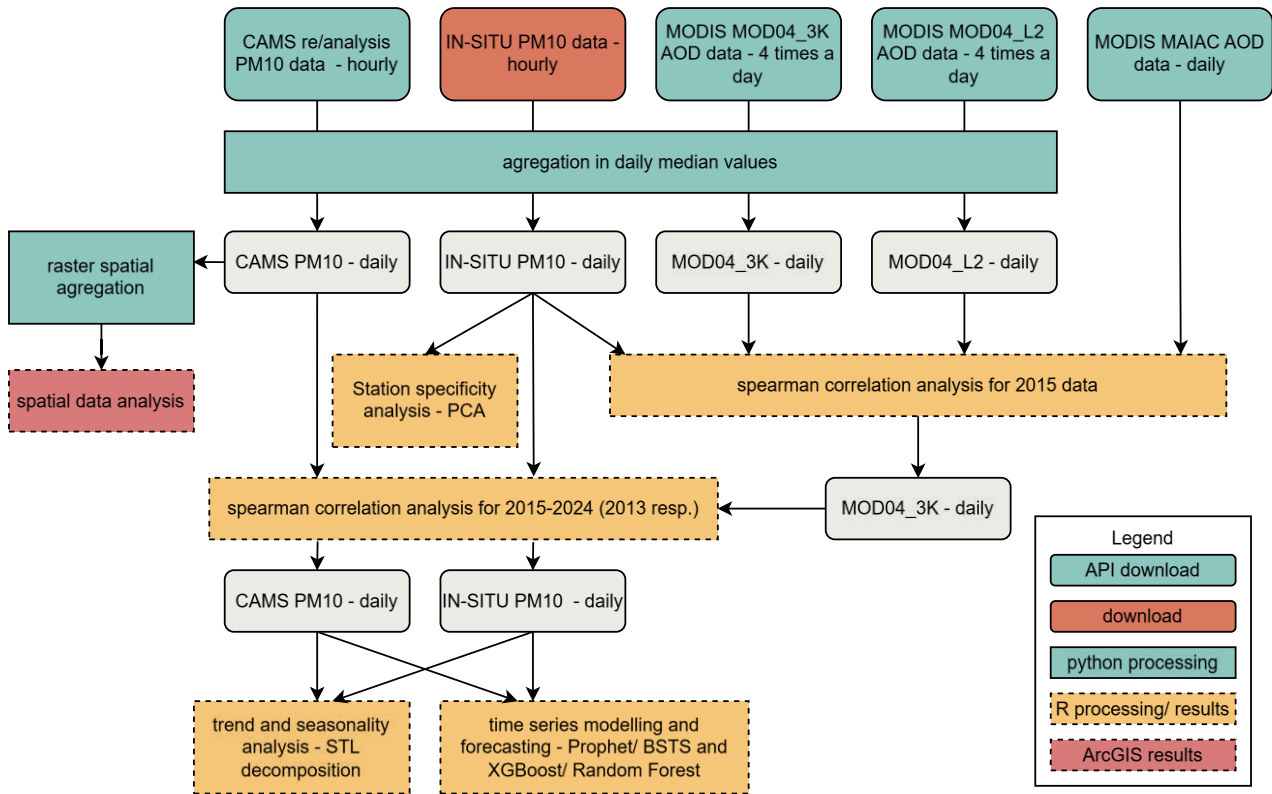


Fig. 2 Visualised processing workflows

2.4.1 Descriptive statistics

First, primary descriptive statistics were used to summarise the distribution and variability of PM₁₀ concentrations at each in-situ monitoring station. Visual tools (histograms and boxplots) and numerical indicators (e.g. completeness, central tendency, spread, and quantiles) provided an initial overview of data quality, typical pollution levels, and the occurrence of extreme values. This step established a baseline understanding of the datasets and subsequent analyses.

2.4.2 Correlation analysis

Because the histograms indicated clear deviations from normality, we used Spearman's rank correlation coefficient (r_s) to quantify the strength and direction of associations between the different datasets. Spearman's correlation is a non-parametric, rank-based measure that assesses the degree of monotonic relationship between two variables, and is therefore suitable for skewed distributions, outliers, and potentially non-linear but monotonic dependencies.

For two variables XX and YY, Spearman's r_s is computed as the Pearson correlation between their ranks, or equivalently (for data without tied ranks):

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)},$$

where d_i is the difference between the ranks of the i -th pair and n is the number of paired observations. The coefficient ranges from -1 to 1 , with $r_s=1$ indicating a perfectly positive monotonic relationship, $r_s=-1$ a perfectly negative monotonic relationship, and $r_s=0$ indicating no monotonic

association. We report both correlation coefficients and associated p-values to assess statistical significance.

Prior to correlation analysis, all time series were temporally harmonised (e.g. aggregated to daily values) and spatially collocated (e.g. using nearest-neighbour or pixel-containing means) to ensure that only synchronous, co-located pairs of observations were compared. Missing values (e.g. due to cloud-contaminated AOD retrievals) were excluded on a pairwise basis and no temporal interpolation was applied, to avoid artificially inflating correlations.

As an initial step, we compared three MODIS AOD products of different native spatial resolutions with in-situ PM_{10} at the Chalkidiki site for the year 2015. This sensitivity test was used to identify which MODIS configuration provided the strongest and most robust association with ground-based measurements (highest r_s , lowest p-value, and reasonable sample size). The best-performing MODIS product was then selected for the full-period correlation analysis, in which we jointly evaluated relationships between MODIS AOD, CAMS PM_{10} , and in-situ PM_{10} over the entire study period and for both study areas.

2.4.3 Multivariate analysis

Principal Component Analysis (PCA) was applied to the in-situ PM_{10} time series to identify common patterns of variability across stations and to assess station specificity. In this context, each station's daily (or monthly) PM_{10} values form a variable in a multivariate dataset, and PCA transforms these correlated station series into a smaller set of uncorrelated components (principal components, PCs). The first few components capture the dominant temporal patterns shared by multiple stations (e.g. a common seasonal cycle or regional background signal), while subsequent components represent more localised or station-specific behaviour (e.g. site-specific peaks related to nearby sources). By examining the loadings of each station on the leading components and visualising scores in the reduced component space, we can group stations with similar pollution dynamics and distinguish them from those with distinct temporal signatures, thereby clarifying to what extent each station reflects local versus regional PM_{10} variability.

2.4.4 Trends and Seasonality analysis

To investigate the long-term temporal dynamics of PM_{10} , we decomposed the time series into trend, seasonal, and residual components using Seasonal-Trend decomposition based on Loess (STL) [38]. STL is an additive decomposition method that iteratively applies locally weighted regression (LOESS) smoothing to separate a time series y_t into:

$$y_t = T_t + S_t + R_t,$$

where T_t denotes the slowly varying long-term trend, S_t the recurring seasonal pattern, and R_t the remaining irregular (noise-like) component.

STL was chosen for the decomposition because it provides a flexible, non-parametric framework in which both the trend and seasonal components are allowed to evolve over time, for example by accommodating changes in the amplitude or shape of the seasonal cycle. The method uses robust LOESS weighting, which reduces the influence of outliers and extreme pollution episodes that are common in air-quality datasets. In addition, STL can handle unequal seasonal amplitudes and is well suited to long time series, making it particularly appropriate for the analysis of PM_{10} time series from atmospheric monitoring and modelling systems.

For this analysis, we applied STL to CAMS surface PM_{10} and in-situ PM_{10} time series. To ensure spatial consistency, in-situ measurements were first aggregated to the CAMS grid using "pixel-containing"

means (i.e. averaging all stations located within each CAMS grid cell). Decomposition was conducted separately for each resulting grid cell (and/or station-equivalent series), using an appropriate seasonal period for daily data (e.g. 365 days) and smoothing windows tuned to capture multi-year trends while preserving the dominant annual cycle.

The extracted trend component T_t was used to assess long-term changes in background PM_{10} levels and to compare large-scale temporal evolution between CAMS and in-situ data. The seasonal component S_t highlighted recurrent intra-annual patterns (e.g. dust episodes and seasonal peaks), enabling a direct comparison of seasonal behaviour across datasets. The residual component R_t was examined to characterise short-term variability and to identify periods of anomalous behaviour potentially linked to episodic dust events or local sources.

2.4.5 Time series modelling and forecasting

Time series modelling and forecasting were applied to complement the retrospective seasonal–trend decomposition and to provide an outlook on future PM_{10} levels.

For medium- to long-term forecasting, two statistical models implemented in R were used: Prophet [39] and Bayesian Structural Time Series (BSTS) [40]. Prophet, an automated procedure developed by Facebook (now Meta), is based on a generalised additive model in which the observed series is expressed as the sum of a piecewise linear or logistic trend, seasonal components (e.g. yearly and weekly), and an error term. It is designed for long, strongly seasonal time series, is robust to missing values and outliers, and requires relatively few manual parameter choices. In parallel, BSTS was applied as a Bayesian state-space framework in which the time series is represented as the sum of latent components such as trend and seasonality, estimated via Monte Carlo Markov Chain (MCMC). Compared with Prophet, BSTS offers greater flexibility in specifying the trend and seasonal structures and can incorporate multiple regressors, but it does not allow missing values in the input series. In this study, Prophet models were implemented using the prophet R package with default seasonal settings and were fitted to daily and monthly PM_{10} time series from CAMS and in-situ data (the latter aggregated to CAMS grid cells where appropriate). BSTS models, implemented with the bsts package, were fitted to monthly CAMS PM_{10} only, using a semi-local linear trend specification, 12 seasonal components (monthly seasonality), and 1000 MCMC iterations after removal of all missing values.

All medium- to long-term models were trained on data up to 31 December 2023 and validated against the subsequent eight months of 2024. Forecasts were then extended to the end of 2025 to obtain a short outlook on possible future PM_{10} trajectories at the study sites. Model performance over the validation period was quantified using root mean square error (RMSE), Pearson correlation coefficient rr between observed and predicted values, and mean absolute error (MAE).

To assess short-term predictive performance, two tree-based machine learning methods—Random Forest [41] and Extreme Gradient Boosting (XGBoost) [42] —were applied to daily CAMS PM_{10} data. Random Forest, implemented via the randomForest package, builds an ensemble of decision trees on bootstrap samples of the data and averages their predictions to reduce variance. In our setup, 500 trees were grown with a minimum terminal node size of 10 observations, and cross-validation was used to evaluate predictive performance. XGBoost, implemented with the xgboost package, is a gradient boosting algorithm in which trees are added sequentially to correct residual errors from the existing ensemble. Here, lagged PM_{10} values (up to 8 previous days) were used as predictors, with a tree-based booster (gbtree), maximum tree depth of 5, and learning rate $\eta=0.1$. In addition to one-day-ahead prediction, an iterative multi-step forecasting scheme was investigated for a 90-day

horizon, in which each newly predicted value was recursively fed back as an input lag for the next day's forecast using the same model configuration. For both Random Forest and XGBoost, the training–validation split matched that of the Prophet and BSTS models (training up to the end of 2023 and validation on the first eight months of 2024), and model performance was again assessed using RMSE, correlation r , and MAE on the validation datasets. These metrics allowed a quantitative comparison of the two approaches and an evaluation of their suitability for PM₁₀ pollution forecasting at the study sites.

2.4.6 Spatial data analysis

Spatial data analysis was conducted to investigate the geographical patterns and evolution of PM₁₀ across the study areas. CAMS outputs were stored as daily median concentration grids in multidimensional monthly files (NetCDF format). For subsequent analysis and visualisation, these daily fields were processed using Python (e.g. `xarray`, `numpy`, and `rasterio` libraries) to compute monthly and annual mean rasters over predefined Areas of Interest (AOIs) at both localities for the entire study period. The aggregation was performed by first subsetting the CAMS grids to the AOI extent and then calculating temporal means on a per-grid-cell basis, while preserving the original spatial resolution and projection of the model data. The resulting monthly and annual composites were used to explore spatio-temporal relationships, identify persistent and emerging PM₁₀ hotspots, and illustrate the temporal evolution of regional dust pollution patterns in a GIS environment.

3 Results

3.1 In-situ local characteristics

The descriptive statistics provide a general overview of the in-situ PM₁₀ concentrations at both study sites and reveal marked differences between them. Both localities exhibit high temporal variability, but overall levels are substantially higher in Greece, where PM₁₀ concentrations range from 0 to 347 $\mu\text{g m}^{-3}$ (mean 22.9 $\mu\text{g m}^{-3}$, $n=30\,496$ values). In contrast, the Finnish site shows a narrower range from 0 to 132 $\mu\text{g m}^{-3}$ (mean 9.90 $\mu\text{g m}^{-3}$, $n=17\,053$). In both regions, the distributions are positively skewed with high kurtosis, indicating right-tailed, heavy-tailed behaviour with frequent moderate and occasional extreme episodes (see Appendix 1). This combination of skewness and kurtosis is typical of air-quality data influenced by episodic events (e.g. dust intrusions, local combustion sources), and suggests that simple Gaussian assumptions are not appropriate for these series.

The PCA applied to the in-situ time series further highlights structural differences between the monitoring networks in the two regions. For the Finnish locality, first principal component (PC1) explained 79,05 % variance and second principal component (PC2) explained 9,21%. Here the station scores are tightly grouped along the first principal component, indicating that most sites share a very similar dominant temporal pattern and differ only weakly in amplitude. By contrast, the Greek stations exhibit a broader spread, PC1 explained 63,39% and PC2 18,17%. Data form several clusters along the PC1, reflecting stronger station-to-station variability in both level and temporal behaviour. In addition, three Greek stations (SAM32, SAM34 and OAM32) clearly deviate from the main group, being separated along PC2 in opposite directions. This separation implies that these stations exhibit distinct temporal signatures (e.g. different seasonal or episodic behaviour) that are not captured by the common regional pattern represented by the first component.

These statistical characteristics are consistent with the visual inspection of monthly means and seasonal cycles at individual stations (see Fig. 3). In Finland, all stations display a broadly coherent seasonal pattern, with peak monthly mean PM₁₀ concentrations of around 15 $\mu\text{g m}^{-3}$ in April and minima of approximately 5 $\mu\text{g m}^{-3}$ during winter months. Inter-annual variability in annual means is

limited, and the highest annual values occur around 2014, after which the long-term trend shows a clear decline, suggesting progressive improvements in air quality and/or favourable meteorological conditions in recent years.

In Greece, despite the smaller spatial extent of the AOI, the dispersion of station values is much larger. Most stations exhibit relatively stable annual means and a modest seasonal cycle characterised by slightly elevated concentrations during the summer, which is consistent with enhanced photochemical activity and contributions from regional dust or resuspended particles. However, three stations (SAM32, SAM33 and SAM34) stand out with annual mean concentrations exceeding those of the remaining sites by up to about $30 \mu\text{g m}^{-3}$. Their seasonal patterns also differ markedly: the coastal station (OAM32) shows maximum values in August, whereas two stations in residential/urban areas (SAM32, SAM34) display pronounced wintertime maxima. Together, these findings underline the stronger local influence and heterogeneity in the Greek network compared with the more regionally homogeneous conditions observed in Finland.

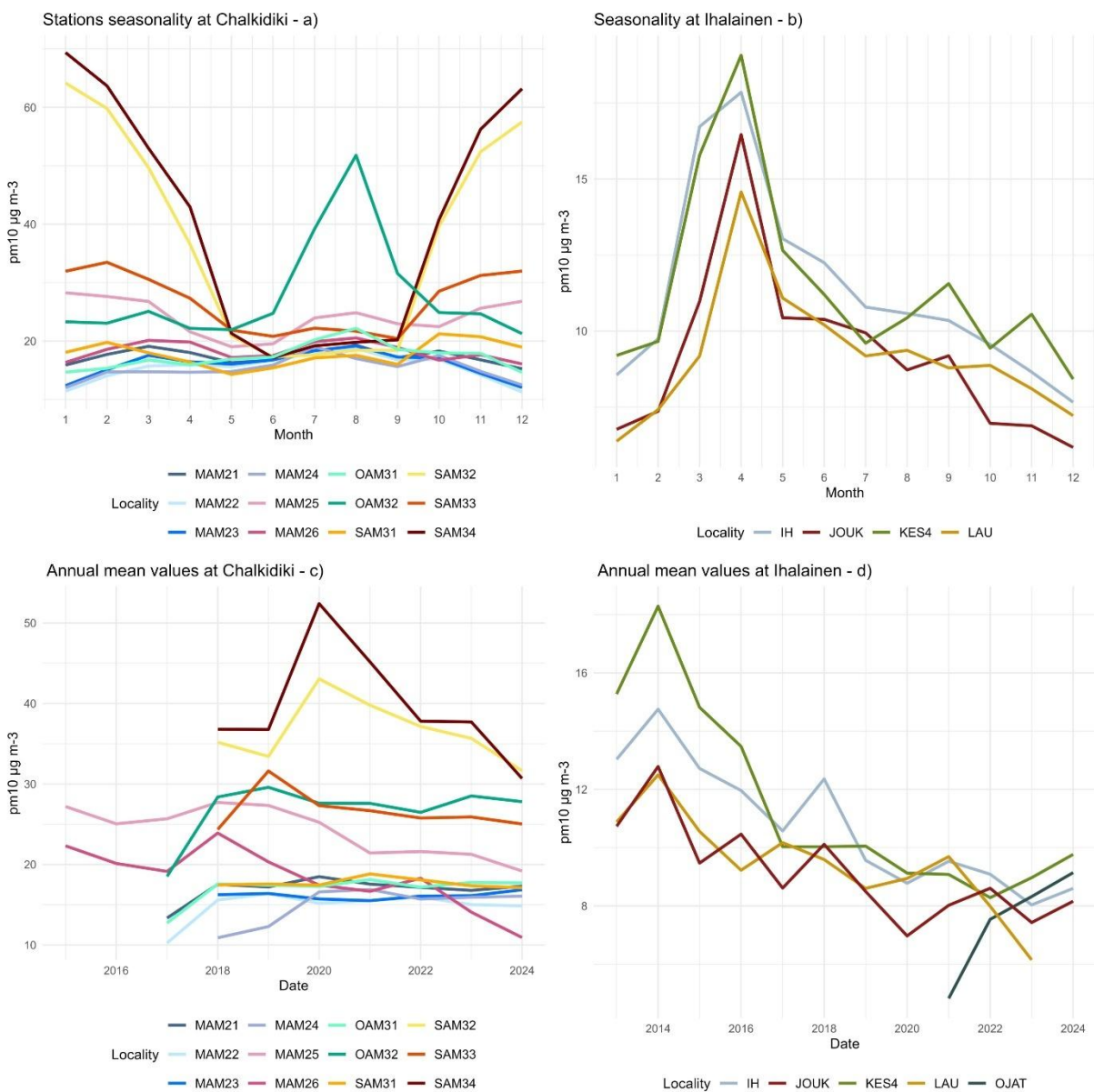


Fig. 3 In-situ data seasonality and trends - Seasonal cycles of PM₁₀ concentrations at a) Chalkidiki and b) Ihalainen locality and Annual mean PM₁₀ concentrations for the monitoring period at c) Chalkidiki and d) Ihalainen

3.2 MODIS AOD and CAMS PM₁₀ products validation

Examining the three MODIS AOD products for 2015 revealed generally weak correlations with in-situ PM₁₀ measurements, largely due to extensive data gaps and retrieval limitations (e.g. clouds, bright surfaces, and quality filtering). Among the products, MOD04_3K (3 km nominal spatial resolution) provided the highest temporal coverage, but still exhibited approximately 61.9% missing values over the Greek study area. Despite this incompleteness, MOD04_3K yielded the strongest correlations with in-situ data in 2015, with Spearman's rank correlation coefficient $r_s \approx 0.35$, compared with $r_s \approx 0.31$ for MOD04_L2 and $r_s \approx 0.18$ for MCD19A2.061. All correlations were computed on temporally matched pairs after discarding missing or invalid AOD and PM₁₀ values. Based on the combination of relatively better data availability and stronger correlations, MOD04_3K was selected and used as supplementary satellite information in the subsequent analyses.

When extended to the full study period, the correlation analysis between in-situ PM₁₀, CAMS PM₁₀, and MODIS AOD indicates pronounced spatial variability, particularly in Greece. At the Greek site (see Fig. 4), Spearman correlations between CAMS and in-situ PM₁₀ range from $r_s \approx 0.3$ at the SAM32 and SAM34 stations to $r_s \approx 0.8$ at MAM21, MAM22, MAM23, and MAM24, suggesting that CAMS reproduces the temporal variability of PM₁₀ substantially better at some locations than others. Correlations between MODIS AOD (MOD04_3K) and in-situ PM₁₀ broadly partition the stations into similar groups but with systematically lower r_s values. At SAM32 and SAM34, a weak negative relationship is observed ($r_s \approx -0.1$), indicating that local conditions or representativeness issues may decouple column AOD from near-surface PM₁₀ at these sites. In contrast, the strongest MODIS–in-situ correlation in Greece is found at MAM22, where $r_s \approx 0.6$, consistent with a more favourable conditions - fewer local emission influences, or better alignment between column and surface aerosol variability.

At the Finnish locality (see Fig. 4), the correlations are more homogeneous and slightly lower overall compared with Greece. For CAMS versus in-situ PM₁₀, r_s values range from approximately 0.37 to 0.70 across stations, while MODIS–in-situ correlations span from about 0.24 to 0.60. Stations KES4, LAU, and OJAT exhibit consistently stronger correlations in both the CAMS and MODIS comparisons, indicating that these locations are more representative of the model grid cell and the satellite footprint, and/or less affected by strong local sources. For all station–dataset pairs at both study areas, the associated p-values are below 0.001, confirming that the observed correlations are highly statistically significant despite the moderate effect sizes and the presence of substantial missing data in the satellite products.

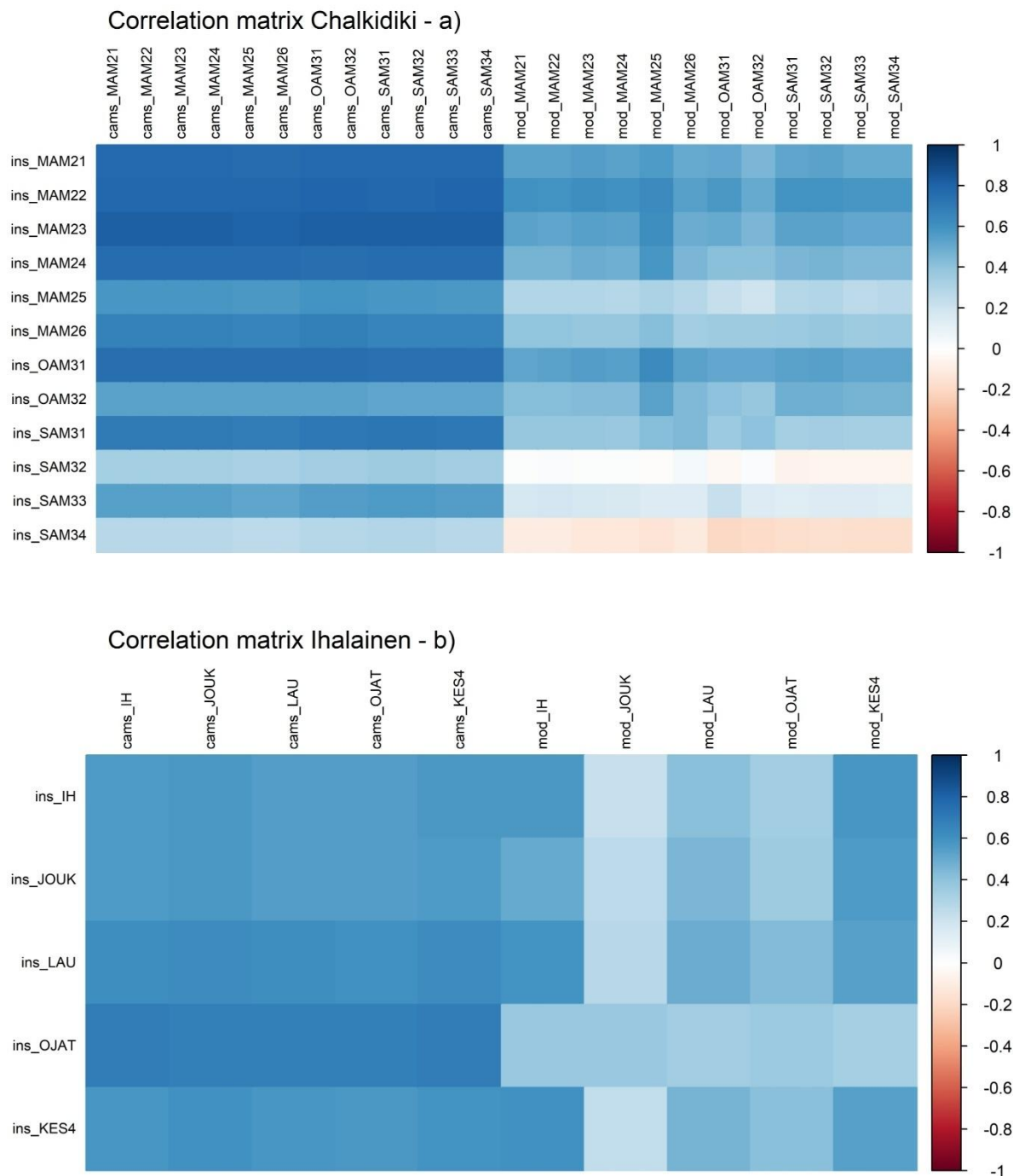


Fig. 4 Correlation results - Correlation matrix of in-situ PM_{10} data, CAMS PM_{10} and MODIS AOD data at both localities, at a) Chalkidiki and b) Ihalainen

3.3 Trends and seasonality

The trends and seasonality of the in-situ data were previously described using monthly and annual mean values. Here, we extend this analysis by directly comparing the STL (Seasonal–Trend decomposition using LOESS) components derived from the *clustered* in-situ time series with those from CAMS for both study areas. For each CAMS grid cell, in-situ measurements from all stations falling within the corresponding 10×10 km pixel were aggregated (pixel-containing mean) to obtain a representative in-situ series at the model resolution, and STL was then applied in an identical way to both datasets (same seasonal period, trend and seasonal smoothing windows).

Across both localities, the CAMS-based PM_{10} values are generally lower, smoother, and more mutually consistent than the in-situ series (see Fig. 5). This behaviour reflects the coarser spatial resolution of CAMS: local peaks and extreme events captured by individual stations are effectively averaged within each 10×10 km grid cell, leading to reduced variance and damped extremes. In the trend component, both CAMS and in-situ data exhibit a broadly comparable long-term evolution, characterised by a predominantly stable to decreasing trend in PM_{10} over the study period at both sites. The agreement between in-situ and CAMS trends is somewhat better in Finland, where local conditions are more homogeneous and station representativeness relative to the model grid appears higher. In Greece, the greater small-scale variability and stronger influence of local sources in case of some stations lead to somewhat larger discrepancies between the in-situ and CAMS trend levels, although the overall direction (gradual decline or stabilisation) remains consistent.

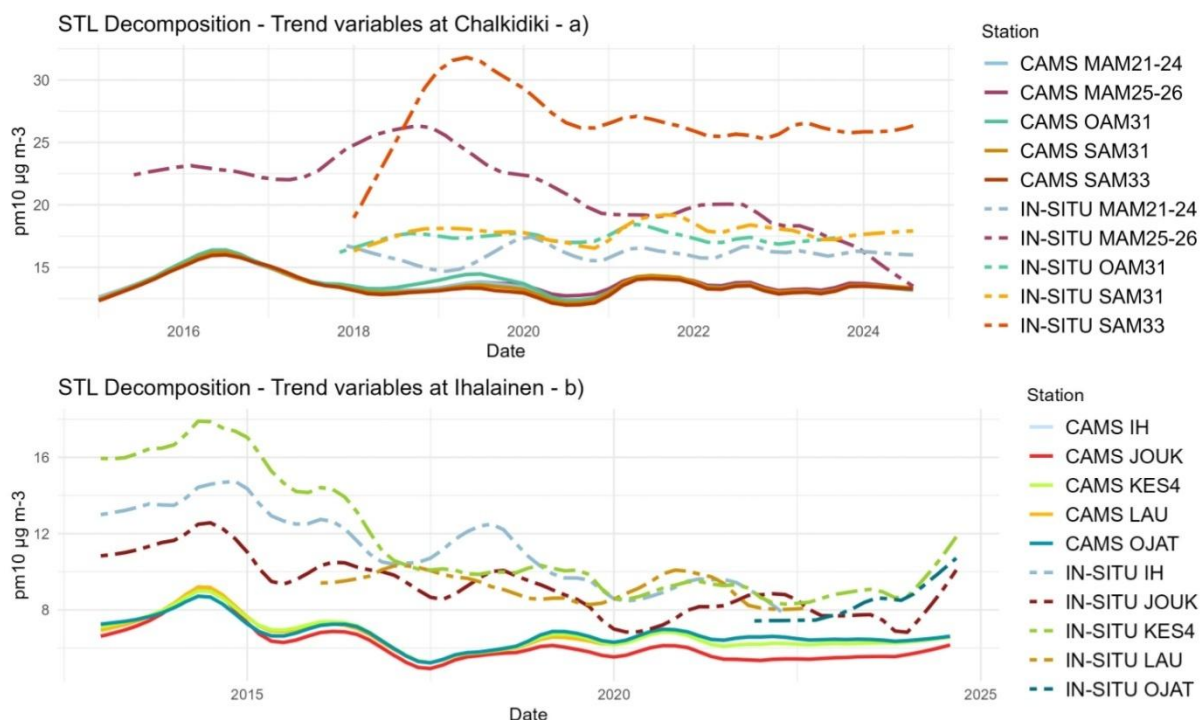


Fig. 5 Trend characteristics – PM_{10} IN-SITU and CAMS data comparison at a) Chalkidiki and b) Ihalainen

The seasonal components from the STL decomposition show systematically higher amplitudes for the in-situ data compared with CAMS, indicating that local stations experience more pronounced seasonal variations than those represented at the grid-cell scale (see Fig. 6). At the Finnish site, the seasonal patterns from both datasets are broadly aligned: both show maxima in early spring, consistent with enhanced resuspension and/or specific meteorological conditions, and minima in autumn and winter. The exact timing of the troughs differs slightly between CAMS and in-situ series, but both place the lowest seasonal values in the colder part of the year.

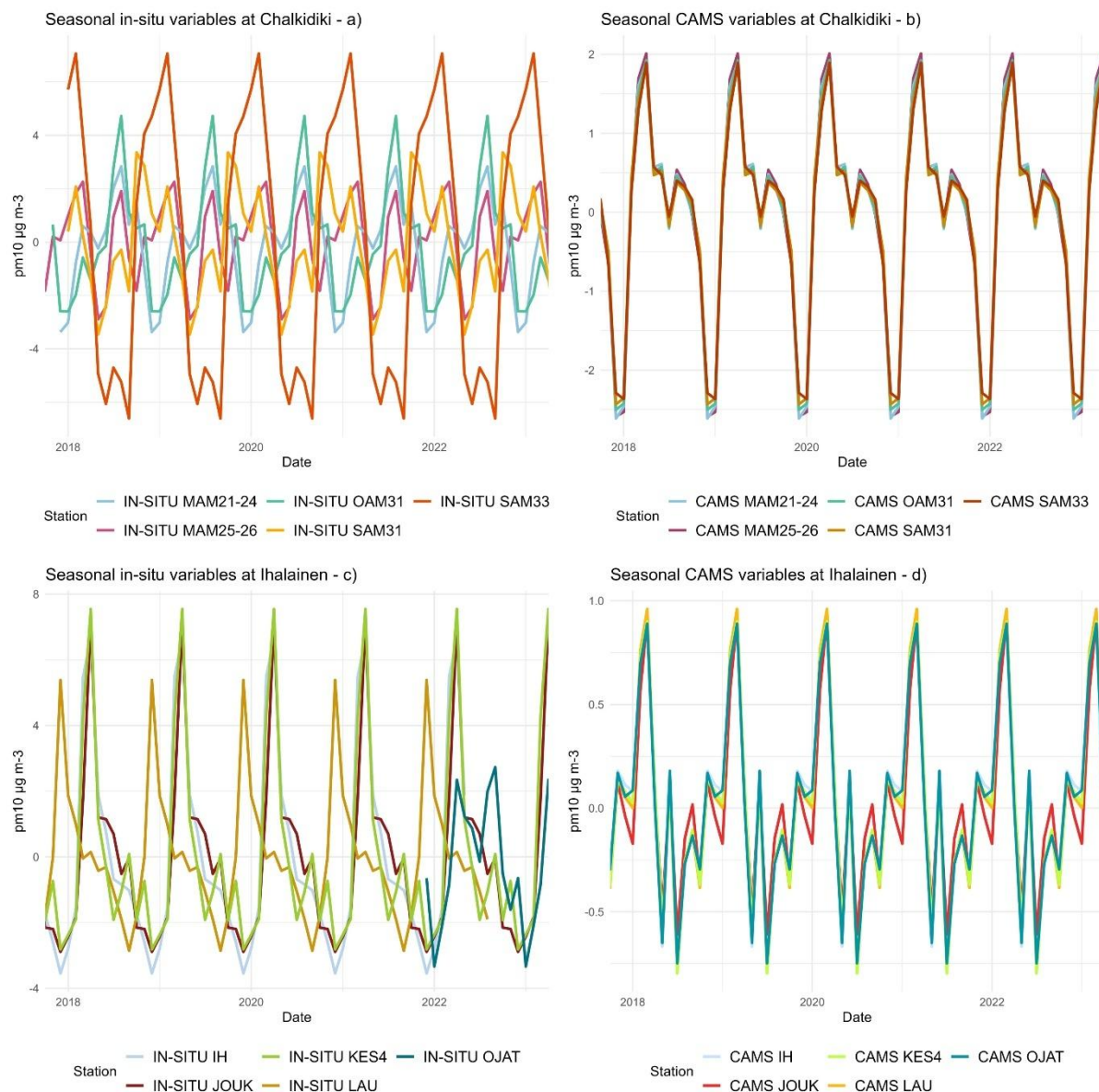


Fig. 6 Seasonal characteristics - comparison of PM_{10} IN-SITU and CAMS data at Chalkidiki (a,b) and Ihalainen (c,d)

3.4 Spatio-temporal patterns

Building upon the previous section, CAMS model fields were used to examine the spatial distribution of PM_{10} across both study areas. For each year, daily CAMS PM_{10} grids were aggregated to annual means and mapped over the AOIs to identify persistent hotspots and regional gradients.

At the Greek site, the highest annual PM_{10} levels were found in 2016. The maps show a clear gradient, with the north-eastern part of the AOI exhibiting the greatest pollution burden, while the south-western monitoring stations consistently record the lowest concentrations. This pattern is consistent with the location of major emission sources and prevailing transport pathways affecting the region (see Fig. 7).

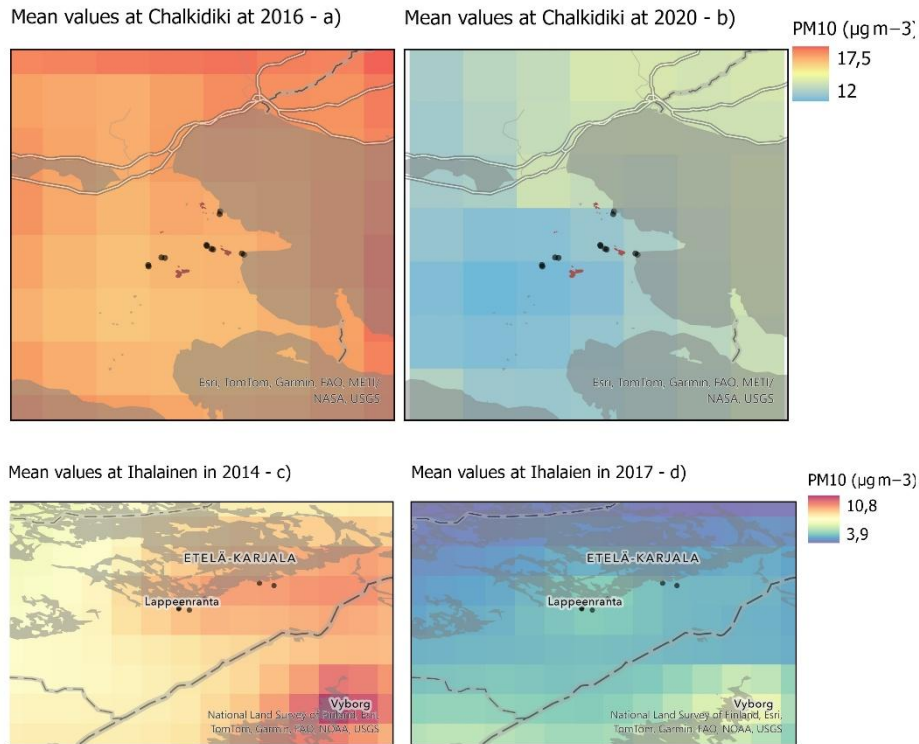
At the Finnish site, the maximum annual PM_{10} values occur in 2014. In contrast to Greece, the most pronounced peaks are centred around the Lappeenranta monitoring stations, reflecting urban and traffic-related emissions. Because the Finnish AOI covers a larger region, an additional hotspot is evident near the town of Vyborg, where modelled PM_{10} concentrations are equal to or even higher

than those near Lappeenranta in some years, indicating a second major source area within the broader domain.

Monthly mean CAMS fields were further analysed to characterise intra-annual variability in spatial patterns. In Greece, the seasonal cycle is relatively smooth, with lower PM₁₀ levels during the winter months (December–January), followed by a rapid increase to a peak in April, likely associated with enhanced regional dust transport and changing meteorological conditions. The spatial pattern also shifts slightly through the year: during winter, higher concentrations are more pronounced in the northern part of the AOI, whereas in summer the highest values tend to occur in the south-eastern corner of the domain (see Fig. 7).

In Finland, the CAMS-derived PM₁₀ distribution is more heterogeneous, reflecting stronger contrasts between urban and natural/remote areas. Elevated monthly means are observed in late winter and early spring (February–March), particularly over urbanised regions such as Lappeenranta and Vyyborg, where road dust resuspension and local emissions are likely to play a major role. The lowest PM₁₀ levels occur at the north-western monitoring stations, where the surrounding environment is less urbanised and background conditions dominate, resulting in consistently lower annual and seasonal concentrations compared with the more densely populated parts of the AOI.

Annual concentrations



Climatological months

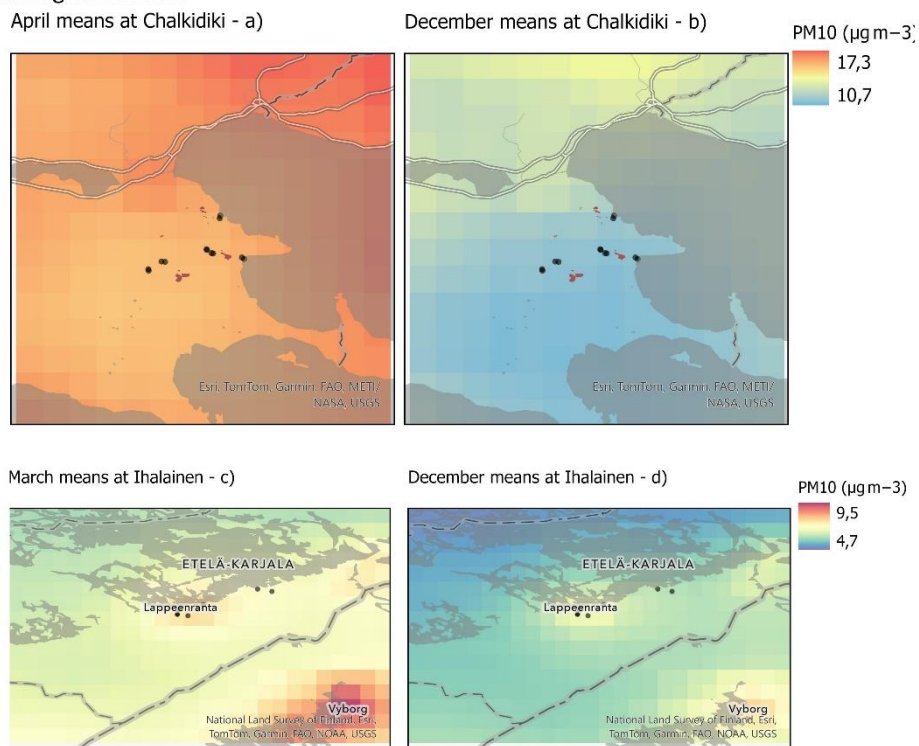


Fig. 7 Annual and climatological visualisations –Maximum and minimum mean annual PM10 pollution from CAMS data at Greek (a,b) and Finland's (c,d) AOI and monthly climatological maxima and minima (monthly mean values thorough temporal range) from CAMS data at Greek (a,b) and Finland's (c,d) AOI in context with local monitoring stations and quarries.

3.5 Time series modelling and forecasting

Long-term forecasts were assessed using the Prophet and BSTS models, both calibrated on data up to the end of 2023 and validated against in-situ observations from January to August 2024. Overall, the two approaches showed comparable performance, with a normalised root mean square error (nRMSE) of approximately 12 %. However, there were regional differences in performance: in Greece, Prophet achieved slightly higher mean correlation with observations ($r \approx 0.32$), whereas in Finland, BSTS and Prophet performed similarly ($r_s \approx 0.17$ vs $r_s \approx 0.18$). Given its ability to natively handle missing values, its lower computational cost, and its fully automated fitting procedure, Prophet was selected as the primary model for subsequent long-term analysis (for performance results at each station see Tab. 2). The resulting trend forecasts point to generally stable or declining PM_{10} levels at most stations, with a more pronounced decrease in Finland, where projected in-situ concentrations decline by roughly 5–10 $\mu g\ m^{-3}$, compared with a weaker decline in Greece.

For short-term forecasting, tree-based machine learning methods were applied to daily data, specifically Random Forest and Extreme Gradient Boosting (XGBoost). Across both regions, XGBoost consistently outperformed Random Forest, yielding higher mean correlations with observed values ($r \approx 0.62$ in Greece and $r \approx 0.47$ in Finland) and lower prediction errors (RMSE and MAE). An iterative multi-step XGBoost setup was also tested to generate forecasts over longer horizons by recursively feeding previous predictions back into the model. Although this iterative approach allowed forecasts to extend further into the future, it led to progressive error accumulation and degradation of performance with lead time, making it unsuitable for robust long-term predictions in this application.

Locality	PROPHET			BSTS			XGBoost			Random Forest			Iterative XGBoost		
Chalkidiki	nRMSE (%)	r	MAE	nRMSE (%)	r	MAE	nRMSE (%)	r	MAE	nRMSE (%)	r	MAE	nRMSE (%)	r	MAE
CAMS MAM21_24	12.2	0.31	1.34	12.3	0.26	1.37	35.1	0.66	3.21	43.1	0.48	4.08	52.8	0.14	5.09
CAMS MAM25_26	11.5	0.33	1.28	11.7	0.26	1.3	37.9	0.6	3.46	43.3	0.51	3.98	51.2	0.16	4.95
CAMS OAM31	12.2	0.31	1.35	12.4	0.33	1.38	36.9	0.61	3.39	42.9	0.49	4.1	52.9	0.1	5.11
CAMS SAM31	13.5	0.29	1.42	13.6	0.33	1.43	37.9	0.61	3.38	43.7	0.5	4.04	54	0.11	5.21
CAMS SAM33	14.2	0.38	1.48	14.3	0.21	1.52	40	0.61	3.51	46.1	0.51	4.17	57.3	0.08	5.42
Chalkidiki mean value	12.7	0.32	1.37	12.9	0.28	1.4	37.6	0.62	3.39	43.8	0.5	4.07	53.6	0.12	5.16
Ihalainen	PROPHET			BSTS			XGBoost			Random Forest			Iterative XGBoost		
Ihalainen	nRMSE (%)	r	MAE	nRMSE (%)	r	MAE	nRMSE (%)	r	MAE	nRMSE (%)	r	MAE	nRMSE (%)	r	MAE
CAMS IH	11.5	0.07	0.56	11.2	0.24	0.61	36.8	0.46	1.76	37.9	0.43	1.91	47.2	0.02	2.35
CAMS JOUK	19.2	0.05	0.99	14.9	0.05	0.78	38.5	0.46	1.7	39.5	0.47	1.83	50.2	-0	2.31
CAMS KES4	12.4	0.31	0.62	10.7	0.19	0.59	36.1	0.48	1.79	37.5	0.46	1.9	47.2	0	2.4
CAMS LAU	9.63	0.26	0.54	10.4	0.17	0.58	35.8	0.47	1.76	37	0.44	1.87	46.4	0.01	2.36
CAMS OJAT	11	0.14	0.61	10.8	0.26	0.61	36.5	0.46	1.75	37.1	0.44	1.87	47.5	0.01	2.39
Ihalainen mean value	12.7	0.17	0.66	11.6	0.18	0.64	36.8	0.47	1.75	37.8	0.45	1.88	47.7	0	2.36

Tab. 2 Results of validation metrics - on long-term Prophet and BSTS CAMS data predictions on short-term XGBoost and Random forest predictions

4 Discussion

The results show that open-pit mining has a measurable influence on local PM₁₀ concentrations, but they also highlight that geographical setting and environmental conditions strongly modulate this impact. Marked differences were observed not only between the two study regions but also among individual stations within each area, reflecting variations in topography, land use, meteorology and proximity to emission sources.

At the Finnish site, PM₁₀ concentrations were generally low and spatially homogeneous, with limited variability between stations and years. The dominant feature was a recurrent spring maximum, with peaks typically occurring in March–April. These peaks likely coincide with the late snowmelt period, when the loss of snow cover exposes accumulated road dust and other particulates, and resuspension of dry material is enhanced by traffic and wind, while surface moisture is temporarily reduced. Frequent precipitation, relatively clean upwind air masses, and a predominantly flat topography favour atmospheric dispersion and wet removal, limiting long-term accumulation of particles. As a result, baseline concentrations remain low and seasonal amplitudes are modest compared with more arid or topographically complex regions.

In contrast, the Greek site exhibited substantially higher PM₁₀ levels and much greater spatio-temporal variability. The semi-arid, hilly terrain is subject to multiple sources and transport pathways. In addition to emissions from mining, traffic and residential heating, the area is regularly affected by long-range dust transport from North Africa, which contributes to elevated PM₁₀ concentrations particularly in spring and early summer. The influence of coastal processes is also evident: coastal stations showed

pronounced summer maxima, likely associated with sea salt aerosols, locally resuspended beach sand and enhanced onshore winds. By contrast, residential and urban stations exhibited winter maxima, driven largely by domestic heating, traffic emissions and more frequent stable boundary-layer conditions that inhibit dispersion. These combined influences lead to a more complex seasonal cycle and sharper spatial gradients than in Finland. Overall, the observed patterns underline how regional meteorology, topography, source mix and land use interact with mining activities to shape PM₁₀ levels and their variability.

The correlation analysis revealed clear differences in the performance and suitability of MODIS AOD and CAMS PM₁₀ datasets for characterising surface particulate pollution in the study areas.

As expected, MODIS showed only weak to moderate correlations with in-situ PM₁₀ (r up to ≈ 0.6 under the most favourable conditions). This relationship was further compromised by a high proportion of missing values ($>50\%$), because AOD retrievals are only available under cloud-free, favourable viewing conditions and are often filtered by strict quality flags (e.g. over bright or heterogeneous surfaces). These issues reduce the temporal representativeness of the satellite record and may bias correlation estimates towards specific meteorological regimes. Such limitations are consistent with previous studies, which indicate that MODIS AOD is best suited to monitoring large-scale dust plumes in arid regions and for climatological analyses, rather than for resolving local air-quality conditions in complex terrain, coastal zones or humid environments, where boundary-layer decoupling and mixed aerosol types blur the link between column AOD and surface PM₁₀.

By contrast, CAMS reanalysis and forecast products showed substantially stronger agreement with in-situ PM₁₀, with Spearman correlations reaching $r_s \approx 0.8$ in Greece and $r_s \approx 0.7$ in Finland. This improved performance reflects the nature of CAMS as a chemistry–transport modelling system constrained by data assimilation of multiple observations (including satellite retrievals, ground-based measurements and meteorological analyses). Previous work has demonstrated that such re/analysis products are effective for capturing regional-scale temporal variability and long-term trends in PM₁₀, particularly where monitoring networks are sparse. However, the coarse spatial resolution of the CAMS products used here (grid spacing of $\sim 10 \times 10$ km) inevitably leads to spatial smoothing: local plumes and sharp gradients near mines, road corridors or dense residential areas are diluted within a grid cell. As a result, local extremes and peak concentrations are systematically underestimated, especially in heterogeneous environments or complex terrain [43]. Despite this limitation, the findings confirm that CAMS is a robust and consistent source of information for regional air-quality assessment, and a valuable complement to in-situ networks, particularly in regions where ground observations are limited or discontinuous [37].

The STL decomposition of both in-situ and CAMS time series revealed broadly downward long-term trends in PM₁₀ at both study sites, indicative of gradual improvements in air quality over the analysis period. In Greece, PM₁₀ levels remained relatively stable or declined slightly at most stations, with the exception of two stations that showed a marked increase around 2020, followed by a continuous decrease thereafter. This behaviour may reflect short-term local changes such as construction works, altered traffic patterns, or specific dust events, superimposed on a longer-term improving background.

In Finland, the downward trend was more pronounced. In-situ data indicated a reduction of approximately $5\text{--}10\ \mu\text{g m}^{-3}$ over the study period, whereas corresponding CAMS estimates suggested a smaller decline of around $1.5\text{--}2\ \mu\text{g m}^{-3}$. This discrepancy suggests that the model may underestimate the magnitude of change, potentially due to limitations in emission inventories, boundary-layer parametrisations, or the spatial averaging intrinsic to the ~ 10 km grid that dampens localised emission reductions (e.g. improved road dust management or stricter emission controls).

Seasonal behaviour also differed between the two regions. In Finland, the STL seasonal components showed consistent peaks in early spring (March–April), associated with snowmelt and enhanced

resuspension of road dust and other deposited particles. The lowest seasonal values occurred in late autumn and winter, when precipitation and snow cover suppress resuspension and long-range transport of mineral dust is minimal. In Greece, seasonal maxima occurred predominantly in late spring and summer, coinciding with dry conditions, strong insolation and frequent Saharan dust intrusions, as well as enhanced resuspension over dry surfaces. CAMS generally reproduced the timing and shape of these seasonal cycles, but with lower amplitudes, again due to spatial averaging and partial smoothing of local extremes within model grid cells.

Based on this study we define several methodological and data-related limitations that should be considered:

Satellite data constraints

- MODIS AOD availability is strongly affected by cloud cover, viewing geometry, surface brightness and quality filtering, which reduces temporal completeness and may introduce sampling biases.
- AOD–PM₁₀ relationships are inherently uncertain in humid, complex or coastal environments, where vertical aerosol profiles, hygroscopic growth and varying aerosol types decouple column optical depth from surface mass concentration.

CAMS spatial resolution and representativeness

- The coarse grid spacing (~10 km) limits the ability of CAMS to capture fine-scale emission hotspots, steep gradients near open-pit mines, road networks or industrial facilities, and micro-scale meteorological effects (e.g. valley circulations).
- Model performance also depends on the accuracy of emissions inventories, boundary-layer dynamics and data assimilation schemes, which may not fully reflect rapid local changes.

In-situ network and representativeness

- Even in relatively well-monitored regions, station density and siting (urban vs rural, background vs roadside vs industrial) influence how representative the measurements are of the surrounding grid cell or satellite footprint.
- Differences in instrument type, maintenance and calibration may introduce additional uncertainty when combining multi-station datasets.

5 Conclusion

The present study delivers a long-term assessment of dust-related PM₁₀ pollution at two sites in Greece and Finland, both characterised by active open-pit mining but subject to markedly different geographical, climatological, industrial and socioeconomic settings. By jointly analysing in-situ measurements, CAMS reanalysis products and MODIS observations, the work highlights how local mining emissions interact with regional background conditions, long-range transport and land-use patterns to shape particulate pollution levels and variability.

Beyond the comparative case study, the analysis establishes a practical framework for integrating optical satellite data (e.g. MODIS AOD) with ground-based monitoring and CAMS reanalysis to support air-quality surveillance and policy applications in mining regions and other remote or hard-to-access areas. In particular, the results confirm that products from the Copernicus Atmosphere Monitoring Service (CAMS) are a valuable resource for regional-scale assessment, providing spatially continuous information over large domains and multi-year periods.

However, the study also demonstrates that, due to their coarse spatial resolution (on the order of 10×10 km) and grid-cell averaging, CAMS fields tend to underestimate local peaks and smooth out fine-scale variability, especially near strong point or line sources (such as mines, industrial facilities and major roads) and in complex terrain. As a result, it is strongly recommended that independent validation data-ideally from well-maintained in-situ networks-be used systematically to evaluate and, where necessary, bias-correct CAMS outputs before they are applied at the local scale for exposure assessment, compliance checking or impact studies.

Taken together, the findings underscore that combining CAMS products with ground-based observations yields a more accurate and robust characterisation of PM₁₀ dynamics across contrasting environmental conditions. This integrated approach enhances the reliability of dust pollution assessments in mining-impacted areas, supports more informed air-quality management and policy decisions, and can be readily transferred to other regions where monitoring infrastructure is limited.

Despite the aforementioned limitations, this study demonstrates that combining ground-based observations with CAMS and MODIS products provides a practical and scalable framework for monitoring and forecasting PM₁₀ in regions affected by open pit mining and other industrial activities. Reanalysis products such as CAMS offer spatially and temporally continuous fields that complement high accuracy but spatially limited in-situ networks. Together with appropriate statistical and machine learning tools, these datasets support evidence-based air quality management, inform mitigation strategies, and enable more robust assessments of long-term trends and population exposure in data scarce regions.

6 Acknowledgements

The presented analysis was conducted under the support of the EU grant MultiMiner. The MultiMiner project is funded by the European Union's Horizon Europe research and innovation actions programme under Grant Agreement No. 10109137474. We would like to thank Ulla Salmela and Hanna Kukkula from Nordkalk for providing valuable information regarding the Ihalainen mining site, and Matthieu Molinier for his helpful comments on the manuscript.

7 Data availability statement

All data used in this study are freely available at addresses listed in chapter 2.2. Data.

8 Competing interests

The authors declare no competing interests.

9 Author contributions

Petra Sedláčková: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Resources, Software, Visualization, Writing

Kostas Gounaris: Data Curation, Writing - Review

Veronika Kopačková - Strnadová: Conceptualization, Funding Acquisition, Project Administration, Supervision, Writing – Review and Editing

656 **10 Reference**

- 657 1. Gaffney, J. S. & Marley, N. A. Atmospheric Chemistry and Air Pollution. *Sci. World J.* **3**, 199–234
658 (2003).
- 659 2. Bai, L., Wang, J., Ma, X. & Lu, H. Air Pollution Forecasts: An Overview. *Int. J. Environ. Res. Public.*
660 *Health* **15**, 780 (2018).
- 661 3. Li, Y. *et al.* Satellite Remote Sensing for Estimating PM_{2.5} and Its Components. *Curr. Pollut. Rep.* **7**,
662 72–87 (2021).
- 663 4. Tiwary, A. & Williams, I. *Air Pollution: Measurement, Modelling and Mitigation*. (CRC Press, Taylor
664 & Francis Group, Boca Raton London New York, 2019).
- 665 5. Gokul, T. *et al.* Particulate pollution and its toxicity to fish: An overview. *Comp. Biochem. Physiol.*
666 *Part C Toxicol. Pharmacol.* **270**, 109646 (2023).
- 667 6. Lippmann, M. Particulate matter (PM) air pollution and health: regulatory and policy implications.
668 *Air Qual. Atmosphere Health* **5**, 237–241 (2012).
- 669 7. Choudhury, A. H., Gordian, M. E. & Morris, S. S. Associations between Respiratory Illness and PM₁₀
670 Air Pollution. *Arch. Environ. Health Int. J.* **52**, 113–117 (1997).
- 671 8. Yu, H. & Zahidi, I. Environmental hazards posed by mine dust, and monitoring method of mine
672 dust pollution using remote sensing technologies: An overview. *Sci. Total Environ.* **864**, 161135
673 (2023).
- 674 9. Abdelkader, M. *et al.* Dust–air pollution dynamics over the eastern Mediterranean. *Atmospheric*
675 *Chem. Phys.* **15**, 9173–9189 (2015).
- 676 10. Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on Ambient
677 Air Quality and Cleaner Air for Europe. *OJ L* vol. 152 (2008).
- 678 11. Burrows, J. P., Platt, U. & Borrell, P. Tropospheric Remote Sensing from Space. in *The Remote*
679 *Sensing of Tropospheric Composition from Space* (eds Burrows, J. P., Borrell, P. & Platt, U.) 1–65
680 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011). doi:10.1007/978-3-642-14791-3_1.

- 681 12. Feuerstein, S. & Schepanski, K. Identification of Dust Sources in a Saharan Dust Hot-Spot and Their
682 Implementation in a Dust-Emission Model. *Remote Sens.* **11**, 4 (2018).
- 683 13. Grgurić, S. *et al.* Relationship between MODIS based Aerosol Optical Depth and PM10 over Croatia.
684 *Cent. Eur. J. Geosci.* **6**, 2–16 (2014).
- 685 14. Randall, V. M. Satellite remote sensing of surface air quality. *Atmos. Environ.* **42**, 7823–7843
686 (2008).
- 687 15. Zhang, Y. *et al.* Satellite remote sensing of atmospheric particulate matter mass concentration:
688 Advances, challenges, and perspectives. *Fundam. Res.* **1**, 240–258 (2021).
- 689 16. El-Harbawi, M. Air quality modelling, simulation, and computational methods: a review. *Environ.*
690 *Rev.* **21**, 149–179 (2013).
- 691 17. Garrigues, S. *et al.* Monitoring multiple satellite aerosol optical depth (AOD) products within the
692 Copernicus Atmosphere Monitoring Service (CAMS) data assimilation system. *Atmospheric Chem.*
693 *Phys.* **22**, 14657–14692 (2022).
- 694 18. Gupta, P., Khan, M. N., da Silva, A. & Patadia, F. MODIS aerosol optical depth observations over
695 urban areas in Pakistan: quantity and quality of the data for air quality monitoring. *Atmospheric*
696 *Pollut. Res.* **4**, 43–52 (2013).
- 697 19. Muhammad Akhlaq, Sheltami, T. R. & Mouftah, H. T. A review of techniques and technologies for
698 sand and dust storm detection. *Rev. Environ. Sci. Biotechnol.* **11**, 305–322 (2012).
- 699 20. Dimitriou, K. & Kassomenos, P. Estimation of North African dust contribution on PM10 episodes
700 at four continental Greek cities. *Ecol. Indic.* **106**, 105530 (2019).
- 701 21. Kukkonen, J., Salmi, T., Saari, H., Konttinen, M. & Kartastenpää, R. Review of urban air quality in
702 Finland. *BOREAL Environ. Res.* **4**: 55-65, (1999).
- 703 22. Höss, A. *et al.* Magmatic and hydrothermal evolution of the Skouries Au-Cu porphyry deposit,
704 northern Greece. *Ore Geol. Rev.* **173**, 106233 (2024).
- 705 23. Voudouris, P., Tarantola, A. & Melfos, V. *A Field Guide to the Tertiary Hydrothermal Ore Deposits*
706 *of Northern Greece.* (2017) doi:10.13140/RG.2.2.16308.45441.

- 707 24. Manigrasso, M. *et al.* PM Dimensional Characterization in an Urban Mediterranean Area: Case
708 Studies on the Separation between Fine and Coarse Atmospheric Aerosol. *Atmosphere* **13**, 227
709 (2022).
- 710 25. Matthaios, V. N., Triantafyllou, A. G. & Koutrakis, P. PM10 episodes in Greece: Local sources versus
711 long-range transport—observations and model simulations. *J. Air Waste Manag. Assoc.* **67**, 105–
712 126 (2017).
- 713 26. Nastos, P. T. Meteorological Patterns Associated with Intense Saharan Dust Outbreaks over Greece
714 in Winter. *Adv. Meteorol.* **2012**, 828301 (2012).
- 715 27. Lapteenranta. *Nordkalk* [https://www.nordkalk.com/about-us/our-](https://www.nordkalk.com/about-us/our-locations/finland/lapteenranta/)
716 [locations/finland/lapteenranta/](https://www.nordkalk.com/about-us/our-locations/finland/lapteenranta/) (2024).
- 717 28. Kukkonen, J. *et al.* Modelling of the public health costs of fine particulate matter and results for
718 Finland in 2015. *Atmospheric Chem. Phys.* **20**, 9371–9391 (2020).
- 719 29. Remer, L. A., Mattoo, S., Levy, R. C. & Munchak, L. A. MODIS 3 km aerosol product: algorithm and
720 global perspective. *Atmospheric Meas. Tech.* **6**, 1829–1844 (2013).
- 721 30. Remer, L. A. *et al.* The MODIS Aerosol Algorithm, Products, and Validation. *J. Atmospheric Sci.* **62**,
722 947–973 (2005).
- 723 31. Sayer, A. M. *et al.* MODIS Collection 6 aerosol products: Comparison between Aqua’s e-Deep Blue,
724 Dark Target, and “merged” data sets, and usage recommendations. *J. Geophys. Res. Atmospheres*
725 **119**, 13,965–13,989 (2014).
- 726 32. EQA reports of global services | Copernicus. [https://atmosphere.copernicus.eu/eqa-reports-](https://atmosphere.copernicus.eu/eqa-reports-global-services)
727 [global-services](https://atmosphere.copernicus.eu/eqa-reports-global-services) (2025).
- 728 33. Eskes, H. *et al.* Technical note: Evaluation of the Copernicus Atmosphere Monitoring Service
729 Cy48R1 upgrade of June 2023. *Atmospheric Chem. Phys.* **24**, 9475–9514 (2024).
- 730 34. Inness, A. *et al.* The CAMS reanalysis of atmospheric composition. *Atmospheric Chem. Phys.* **19**,
731 3515–3556 (2019).
- 732 35. Marécal, V. *et al.* A regional air quality forecasting system over Europe: the MACC-II daily ensemble
733 production. *Geosci. Model Dev.* **8**, 2777–2813 (2015).

36. Copernicus. <https://atmosphere.copernicus.eu/about-us> (2024).
37. Gualtieri, G. *et al.* Assessing capability of Copernicus Atmosphere Monitoring Service to forecast PM_{2.5} and PM₁₀ hourly concentrations in a European air quality hotspot. *Atmospheric Pollut. Res.* **16**, 102567 (2025).
38. Cleveland, R. B., Cleveland, W. S., McRae, J. E. & Terpenning, I. STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *J. Off. Stat.* **6**, (1990).
39. Taylor, S. & Letham, B. prophet: Automatic Forecasting Procedure. 1.0 <https://doi.org/10.32614/CRAN.package.prophet> (2017).
40. Scott, S. L. bsts: Bayesian Structural Time Series. 0.9.11 <https://doi.org/10.32614/CRAN.package.bsts> (2014).
41. Breiman, L., Cutler, A., Liaw, A. & Wiener, M. randomForest: Breiman and Cutlers Random Forests for Classification and Regression. 4.7–1.2 <https://doi.org/10.32614/CRAN.package.randomForest> (2002).
42. XGBoost Documentation. *XGBoost Documentation — xgboost 2.1.3 documentation* <https://xgboost.readthedocs.io/en/stable/> (2022).
43. Velea, L., Udristioiu, M. T., Bojariu, R. & Sararu, S. Statistical characteristics of particulate matter (PM₁₀) concentration in Romanian selected urban areas based on CAMS-regional ensemble model reanalysis. *AIP Conf. Proc.* **2218**, 030003 (2020).

Appendix 1

Histograms and PCA analysis results of In-situ data at both localities in Greece (a) and Finland (b)

