Environmental Research 298 (2026) 124273

Contents lists available at ScienceDirect

Environmental Research

journal homepage: www.elsevier.com/locate/envres

An ensemble-forecasting model for airborne grass pollen at three climatically distinct sites

Maria P. Plaza [a,b,*], Jose Oteros [c,d], Vivien Leier-Wirtz [a], Athanasios Charalampopoulos [e], Carmen Galán [c,d], Caroline Holzmann [a], Franziska Kolek [a], Despoina Vokou [e], Claudia Traidl-Hoffmann [a,b,f], Stefanie Gilles [a,b], Athanasios Damialis [e]

[a] Institute of Environmental Medicine and Integrative Health - Environmental Medicine, Faculty of Medicine, University of Augsburg and University Hospital of Augsburg, Augsburg, Germany
[b] Institute of Environmental Medicine, Environmental Health Center, Helmholtz Zentrum München, Neuherberg, Germany
[c] Department of Botany, Ecology and Plant Physiology, Agrifood Campus of International Excellence (CeiA3), University of Cordoba, Spain
[d] Andalusia Inter-University Institute for Earth System Research (IISTA), University of Cordoba, Rabanales Campus, Celestino Mutis Building, E-14071, Córdoba, Spain
[e] Aerobiology Unit, Department of Ecology, School of Biology, Faculty of Sciences, Aristotle University of Thessaloniki, 546 36, Thessaloniki, Greece
[f] Christine Kühne - Center for Allergy Research and Education (CK-CARE), Davos, Switzerland

ARTICLE INFO

Keywords:
Aerobiology
Ecological informatics
Ecological modelling
Environmental forecasting
Meteorological predictors
Public health

ABSTRACT

Precise airborne pollen forecasting is essential for mitigating exposure risks in individuals with pollen-related respiratory diseases such as allergic rhinitis and asthma and for supporting timely public health warning. Moreover, long-term accurate pollen forecasts could also support biodiversity conservation, ecosystem functions, and public-health protection.

We developed an ensemble forecasting model for airborne grass (Poaceae) pollen concentrations in three climatically distinct European cities: Augsburg (Germany, transitional temperate-continental), Córdoba (Spain, dry Mediterranean), and Thessaloniki (Greece, humid Mediterranean). Pollen data (2018-2024) from Hirst-type volumetric traps were combined with meteorological parameters (temperature, humidity, precipitation). The 2024 pollen data were used for validation. Of 61 candidates, seven representative model families (Regularized Linear Regression, Extreme Gradient Boosting, Neural Network Autoregression [NNETAR], Random Forest, Support Vector Regression, Prophet–XGBoost hybrid, and Autoregressive Integrated Moving Average [ARIMA]) were selected for the ensemble. Model weights were assigned according to predictive performance.

The ensemble achieved R^2 values of 0.66 in Augsburg, 0.62 in Córdoba and 0.84 in Thessaloniki, with NNETAR and/or ARIMA contributing most strongly during the pollen season. Lagged pollen concentrations and previous-day temperature emerged as key predictors. When incorporating data from an automatic pollen monitor (BAA500, Helmut Hund GmbH) in Augsburg, the model achieved higher predictive performance ($R^2 = 0.89$). Our findings demonstrate that ensemble-based pollen forecasting may generalize across contrasting bioclimatic regions, while remaining sensitive to local ecological and climatic controls. This framework provides a foundation for more powerful (real-time) forecasting systems aimed primarily at improving daily allergy risk management, while potentially offering complementary insights into longer-term vegetation dynamics under climate variability.

1. Introduction

Monitoring of airborne pollen and other biological air particles has been conducted for almost a century now, placing the associated scientific field - Aerobiology - at the intersection of biological, medical and environmental sciences. Complex ecological systems, numerous organisms from various taxonomic groups, and exploration and interpretation of avowedly neglected ecosystems on the planet and of not fully known biodiversity set airborne particles of biological origin to the spotlight of the Intergovernmental Panel on Climate Change, being characterised as

* Corresponding author. Environmental Medicine - Institute of Environmental Medicine and Integrative Health, Faculty of Medicine, University of Augsburg and University Hospital of Augsburg, Stenglinstrasse 2, 86156, Augsburg, Germany.
E-mail address: maria.plaza@med.uni-augsburg.de (M.P. Plaza).

https://doi.org/10.1016/j.envres.2026.124273
Received 30 January 2026; Received in revised form 10 March 2026; Accepted 12 March 2026
Available online 14 March 2026

climate change bio-indicators in the latest WGII AR6 (IPCC, 2022). In recent years, interest in aerobiology has grown noticeably. Many of the challenges driven by climate change across plant and microbial ecology, urban ecology, phytopathology, agriculture, and forestry have highlighted the need for reliable aerobiological monitoring. This type of monitoring supports a broad range of ecological questions, from flowering and fruit phenology to the links between flower and seed or fruit production (Oteros et al., 2014), and can provide indirect insights into vegetation dynamics under changing environmental conditions. However, daily pollen forecasting is primarily motivated by applications in exposure assessment and risk mitigation rather than direct inference of reproductive processes such as cross-pollination, pollinator decline or gene flow. As a result, biomonitoring and related forecasting have become increasingly important tools, particularly for health-related and environmental monitoring purposes. In addition, they provide timely information that can help meet ecological objectives, make better use of ecosystem services, and reduce risks associated with a changing climate.

Increasingly, it is also recognised that airborne pollen dynamics are not controlled by climate alone, but are strongly shaped by vegetation structure, land-use patterns, and landscape composition. Recent work combining biological observations with remote sensing has demonstrated that grass pollen emission is tightly linked to surrounding land cover and habitat configuration, highlighting the importance of ecological context for understanding and forecasting pollen release (Rojo et al., 2022).

This need becomes even more evident when we also consider the impact of certain airborne biological particles, most notably pollen, on human health. These particles are well known to contribute to respiratory conditions such as allergic rhinitis and asthma (Annesi-Maesano et al., 2023; Pfaar et al., 2020), creating a substantial public-health burden worldwide. Their prevalence is projected to increase under the combined influence of climate change (Gehrig and Clot, 2021), urbanization and shifting land-use patterns (e.g. Oh, 2022). As allergy has been characterised an environmental disease and, because of its inter- and multi-disciplinary nature, it has attracted the attention of ecologists worldwide for many years (e.g. Damialis et al., 2019 and references therein). Among aeroallergens, airborne grass (Poaceae) pollen stands out as one of the most prevalent and potent triggers of allergic symptoms throughout the globe, making accurate forecasting of its concentrations and dynamics essential for early warning systems, health advisories, and ecological monitoring (García-Mozo, 2017).

In the light of this, the tools used to monitor airborne pollen become critically important. Because pollen emissions integrate both biological production and landscape structure, monitoring systems also act as indirect sensors of vegetation functioning and land-use patterns at the ecosystem scale (López-Orozco et al., 2023). Historically, volumetric monitoring devices, primarily Hirst-type traps, have served as the reference standard for airborne pollen and spore sampling, providing reliable long-term datasets via microscopy-based taxonomy (Hirst, 1952). However, the required manual classification and counting effort introduces delays of several days, along with potential inconsistencies in taxonomic identification due to expert's subjectivity (Milic et al., 2020; Smith et al., 2019). To address these limitations, automated pollen-monitoring technologies that combine optical detection with machine learning classification are increasingly adopted (Buters et al., 2024). Recent studies show that such devices, including systems like the KH3000, SwisensPoleno and BAA500, correlate well with Hirst-type pollen data concentrations, although systematic biases, particularly during peak periods, persist (Liu et al., 2025; Maya-Manzano et al., 2023; Plaza et al., 2022).

Forecasting approaches naturally follow monitoring, as they increase the value of observations by using historical records and environmental variables, such as meteorological data, to anticipate future concentrations. From a public-health perspective, forecasting enables proactive exposure management and timely risk communication. Precise forecasts are essential for effective clinical allergy management and the operative functioning of public health early warning systems (Holzmann et al., 2025; Zhu et al., 2024). From an ecological perspective, forecasting additionally allows the anticipation of flowering pulses, dispersal windows and regional pollen transport under changing environmental conditions (Kurganskiy et al., 2021). Grass pollen forecasting is also increasingly relevant for tracking the ecological consequences of urban greening (Stevanovic et al., 2025), agricultural intensification (Devadas et al., 2018), and the indirect influence of biological invasions that reshape allergenic species assemblages and seasonal exposure patterns across Europe (Galán Díaz et al., 2024).

Early attempts relied on empirical and mechanistic approaches, often adapted from atmospheric dispersion models originally developed for air pollution studies (e.g., Siljamo et al., 2013). While such models can simulate transport and deposition, they are limited in their ability to capture the biological drivers of flowering and pollen release. With the development of environmental monitoring networks and access to larger datasets, statistical and machine learning methods were introduced, ranging from multiple linear regression to neural networks (Rojo and Pérez-Badia, 2015). These approaches improve short-term predictive performance by exploiting correlations between pollen concentrations and meteorological predictors, lagged pollen records, phenological indices, and land-surface variables (Cebrino et al., 2016; Cordero et al., 2021; Lo et al., 2021; Zewdie et al., 2019) often yielding significant correlation coefficients on independent test data. Recent studies emphasize that the integration of machine learning with environmental and biomedical datasets can substantially improve prediction accuracy, particularly when deep learning and hybrid models are applied to complex, nonlinear systems such as pollen dynamics (Silver et al., 2020). Nonetheless, many remain limited, relying either on purely statistical relationships or on physical dispersion assumptions without integrating both perspectives.

Data-driven studies, such as Voukantsis et al. (2010) in Thessaloniki, demonstrated that statistical and machine learning methods can provide reliable daily pollen predictions. More recently, deep learning methods have advanced this field: Picornell et al. (2024) showed that LSTM networks can effectively capture annual pollen dynamics, while Nickovic et al. (2023) emphasized the role of forecasting airborne pollen and sub-pollen particles in assessing thunderstorm asthma risks. Together, these studies highlight the growing importance of hybrid and machine learning models in operational pollen forecasting. More recently (Zhu et al., 2024), reviewed forecasting approaches, highlighting the need for improved interpretability, generalizability across climates and taxa, and assimilation of real-time data streams.

Parallel to empirical and machine learning methods, numerical pollen dispersion models (e.g., COSMO-ART, SILAM) provide spatially explicit forecasts of pollen emission, transport, deposition, and transformation processes (Adamov and Pauling, 2023; Hernandez-Ceballos et al., 2014; Hernández-Ceballos et al., 2026). Yet, the integration of observed pollen data remains limited by delays in measurement availability. Adamov and Pauling (2023) proposed a calibration-based assimilation method to better leverage real-time observations over extended forecast horizons.

Alongside these advances, ensemble learning has emerged as a powerful paradigm in predictive modelling. By aggregating multiple base learners, ensembles achieve higher accuracy and stability than individual models (Zhu et al., 2024). Studies report that weighted ensembles, hybrid models, and data fusion approaches confer greater forecast stability compared with single algorithms. Recent aerobiological studies further support this trend, showing that ensemble and hybrid machine-learning approaches can improve both spatial pollen prediction and short-term forecast robustness under variable environmental conditions (Valipour Shokouhi et al., 2024b; Yin et al., 2026). For example, a study in Beijing demonstrated that a weighted ensemble reached one-day coefficients of determination above 0.67, while tree-based models such as Extreme Gradient Boosting performed slightly higher individually (Ruan et al., 2024). Investigations in Europe show

that data-fusion methods outperform simple averages or medians, maintaining nearly optimal performance up to four days ahead (Cordero et al., 2021; Sofiev et al., 2015, 2017). Similar adaptability of ensemble approaches has been demonstrated across regions and taxa, including ragweed, birch, olive, and grass pollen (Liu et al., 2017; Zewdie et al., 2019).

Ensemble modelling has also been explored in pollen dispersion: for instance, the MACC multi-model birch pollen ensemble was one of the first European efforts to combine outputs from several dispersion models, improving phenology reproduction though still struggling with absolute concentration accuracy (Sofiev et al., 2015). Nevertheless, short-term forecasts continue to face challenges due to multiple sources of variability: (i) inter-annual fluctuations in flower timing and pollen production, which depend on climatic conditions during both the current and previous year and on internal physiological processes of the plant (Rojo et al., 2015); (ii) local geographic influences, including topography, altitude, and microclimate that shape flowering phenology (Aguilera and Valenzuela, 2012; Martínez-Bracero et al., 2025; Oteros et al., 2013; Rojo and Pérez-Badia, 2015); and (iii) high variability in maximum pollen concentrations between years and even between consecutive days (Fernández-Rodríguez et al., 2016). These complexities mean that no single model is sufficient across all sites and seasons.

Taken together, these advances underscore the need for forecasting approaches that can keep up with both, ecological complexity (e.g. climatic forcing, land-use changes and intensity and vegetation management) and the public-health relevance of airborne pollen. Although many modelling approaches now exist, ranging from statistical and machine-learning methods to numerical dispersion models, they remain difficult to apply consistently across different regions, climates, or monitoring systems. In this study, we present an ensemble-based framework for forecasting airborne grass pollen in three contrasting European environments, integrating statistical, tree-based, and neural-network models and evaluated them using data from both automatic and conventional instruments. With this, we aim to see whether combining different modelling strategies can improve the robustness and transferability of pollen forecasts. While ensemble learning is well established in machine learning and increasingly explored in aerobiological forecasting, comparative evaluations across climatically contrasting sites, harmonized predictors, and different monitoring system remain limited, and this study addresses that operational gap. Our work offers a practical step toward more reliable forecasting tools that can primarily support allergy-related early-warning systems, while also providing complementary, indirect insights into vegetation responses to contrasting climatic and land-use settings at regional scales.

2. Materials and methods

2.1. Study sites and data sources

The study was conducted at three geographically and climatically distinct sites: Augsburg in southern Germany, Córdoba in southern Spain, and Thessaloniki in northern Greece (Fig. 1). Augsburg is characterized by a relatively cool and humid climate with rich vegetation that includes mixed forests and heathlands, partly due to proximity to river floodplains and surrounding woodland. By contrast, Córdoba has a hotter and much drier Mediterranean climate, with extremely high summer temperatures, low annual precipitation, and very limited rainfall during the dry season. The vegetation is dominated by drought-resistant species and Mediterranean scrubland, with extensive agricultural landscapes in the surrounding area, including olive groves and cereal fields. Thessaloniki, belonging also to a Mediterranean-type climate, experiences hot, dry summers but milder, more humid winters than Córdoba, with a complex topography influenced by its coastal location on the Thermaic Gulf and the surrounding mountainous terrain. The vegetation reflects this more temperate setting, combining Mediterranean shrubland with deciduous woodland and diverse agricultural areas. These differences in climate and vegetation shape distinct pollen source compositions and seasonal patterns, providing valuable insights into atmospheric pollen dynamics for evaluating the ensemble-forecasting framework.

Airborne Poaceae pollen monitoring was carried out with the traditional volumetric Hirst-type samplers (Hirst, 1952; Burkard Co. Ltd, UK) at all three study sites. In Augsburg the device was located at the Institute of Environmental Medicine and Integrative Health (48°23′02.6″N, 10°50′36.3″E; 494 m a.s.l.), while in Córdoba the Hirst trap was positioned at the University Campus of Rabanales (37°53′0″N,

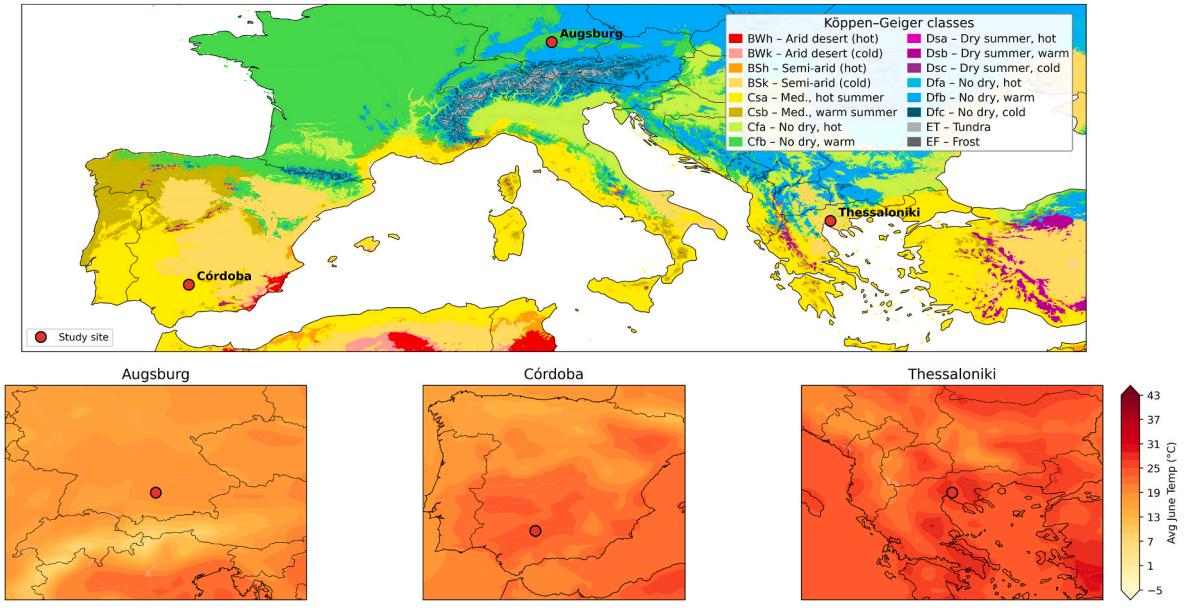

Fig. 1. Climate classification and early-summer thermal context of the study sites. Köppen-Geiger climate classification (Beck et al., 2023) across Europe (top panel) and corresponding regional distributions of mean 2-m air temperature for June 2024 (bottom panels) centered on each study site (Augsburg, Córdoba and Thessaloniki). The Köppen-Geiger map represents the 1991-2020 climatological baseline, whereas the zoom maps illustrate the 2024 early-summer thermal conditions at finer spatial scale based on ERA5 reanalysis data.

4°48′0″W; 123 m a.s.l.). In Thessaloniki, the sampler was installed at the Aristotle University Campus (40°37′59″N, 22°57′35″E; 490 m a.s.l.), providing representative measurements of airborne pollen in this urban-coastal Mediterranean setting. In addition to the Hirst sampler, Augsburg was equipped with an automatic, near-real-time pollen monitoring device (PoMo - BAA500, Helmut Hund GmbH, Wetzlar, Germany), which was co-located with the Hirst-type trap at the same site. The traditional Hirst sampler provides a well-established reference method based on continuous volumetric sampling onto adhesive-coated tapes, followed by microscopic identification and counting of pollen grains. PoMo, in contrast, combines automatic image recognition with classification algorithms to provide near-real-time data. The co-location of both instruments in Augsburg enables a direct comparison of traditional manual counting with automated pollen detection, while maintaining consistency with the reference standard used in Córdoba and Thessaloniki. For the Hirst-type volumetric traps, raw pollen counts were provided as bi-hourly observations, whereas the PoMo device reported three-hourly measurements; in both cases, observations were aggregated and converted into daily mean pollen concentrations (pollen grains/m^3) according to established aerobiological guidelines, ensuring comparability across stations and device types.

Prior to the seasonal analysis, quality controls were performed on all pollen time series to confirm data continuity and identify missing observations in each station. Small gaps of typically 1-3 consecutive days were filled using the *tseries* method in the AeRobiology package (Rojo et al., 2019), which employs an STL-based interpolation method (seasonal trend decomposition by LOESS; Cleveland et al., 1990). No long gaps or station shutdowns were identified. Season parameters were determined using the 95% cumulative method (Andersen, 1991), with the season start defined as the date when cumulative pollen reached 2.5% of the annual total and the end when it exceeded 97.5%.

Daily meteorological parameters, including average relative humidity (% RH), maximum, minimum and average temperature (°C), as well as precipitation (mm), were provided by the CDC (Climate Data Center) of the German Meteorological Service (DWD) for Augsburg, by the Spanish State Meteorological Agency (AEMET) for Córdoba, and by the Hellenic National Meteorological Service (HNMS) for Thessaloniki.

2.2. Data preprocessing

Pollen and meteorological datasets were merged into a unified daily time series for each monitoring station. To account for short-term temporal dependence and smoothed persistence, two lagged predictors were created from the daily Poaceae pollen concentration series of each station s ($P_{s,t}$): a one-day lag ($P_{s,t-1}$) and a three-day moving average of the lagged series ($\overline{P}^{(3)}_{s,t-1}$), computed as:

$$\overline{P}^{(3)}_{s,t-1} = \frac{1}{3}\sum_{i=1}^{3} P_{s,t-i}$$

We also derived a binary presence indicator (*PollenYN*) from the lagged value, defined as "yes" if $P_{s,t-1} > 0$ and "no" if otherwise. Meteorological predictors were included in the model as continuous variables (mean temperature ($T_{s,t}$ in [^{0}C]), relative humidity ($RH_{s,t}$ in [%]) and precipitation ($R_{s,t}$ in [mm])). Precipitation was additionally encoded as a categorical variable $RainLevels_{s,t}$ with four levels (No Rain: $0 \leq 1$ mm; Low: $>1 \leq 10$ mm; Average: $>10 \leq 50$ mm; Heavy: >50 mm) to reflect rainfall-intensity thresholds commonly used in climatological classification (World Meteorological Organization, 2017). While precise thresholds vary by region, our cut-points align with the range of values used in international guidelines and enable meaningful differentiation of rainfall regimes in the pollen-forecasting context.

Seasonal structure was captured by extracting time-series signatures from the calendar date (week-of-year, month, and year). From these, high-frequency components (day/hour/minute) were excluded. Continuous temporal indices (e.g., numeric date, year) were standardized using z-score normalization (mean = 0, standard deviation = 1), while primary meteorological and pollen-derived predictors (e.g., mean temperature, relative humidity, precipitation, 1-lag and 3-day moving average) were retained on their original scales. Categorical variables (*PollenYN*, rain levels, and date-part factors) were dummy-encoded. Rows with missing values (including those introduced by lags/moving averages) were dropped prior to model training. For each station s and day t, the response is the daily airborne Poaceae pollen concentration $x_{s,t}$, in pollen grains/m^3, modeled on its original (untransformed) scale.

2.3. Forecasting framework and sub-models

A total of 61 candidate models were screened using the *caret* package (Kuhn, 2008) in *Rstudio* (Posit team, 2023; R Core Team, 2022). The dataset was divided into a training period (2018-2023) and an independent validation year (2024). After preprocessing, including the exclusion of days for which all modeled variables (pollen and meteorological predictors) were equal to zero and therefore not informative for model fitting, the training datasets comprised 2191 and 1988 daily observations for Augsburg (Hirst-type sampler and automatic PoMo monitor, respectively), 2005 for Córdoba, and 2190 for Thessaloniki, while the corresponding validation datasets consisted of 366, 268, 306, and 365 observations, respectively. All models were trained using rolling-origin resampling to respect temporal dependence. To formalize the model inputs, we defined for each station s and day t the predictor vector:

$$x_{s,t} = T_{s,t}, RH_{s,t}, R_{s,t}, P_{s,t-1}, \overline{P}^{(3)}_{s,t-1}, PollenYN_{s,t}, RainLevels_{s,t}, d^{(time)}_{s,t}$$

where $d^{(time)}_{s,t}$ represents the engineered time-signature features (e.g., week-of-year, month, numeric time index).

Comparative performance was evaluated using four standard applied forecasting power metrics: the coefficient of determination (R^2), root mean squared error (RMSE), mean absolute error (MAE) and mean absolute scaled error (MASE). RMSE and MAE quantify absolute prediction error, while R^2 measures explained variance. MASE was calculated relative to a one-step naïve persistence forecast ($\hat{y}_t = y_{t-1}$), providing a scale-independent measure of performance relative to short-term autocorrelation in the pollen time series.

The initial screening was conducted on the Augsburg (Hirst) dataset as the development site using the independent 2024 validation year for strict out-of-sample comparison. Models were first ranked according to their RMSE performance, and those within the top performance quartile were identified as the competitive tier. From this subset, representative algorithms were retained according to three complementary criteria: (i) competitive predictive performance on the independent test set (models within the top quartile of predictive performance across RMSE and R^2), (ii) methodological diversity (linear, tree-based, kernel-based, neural, hybrid time-series), and (iii) structural complementarity in terms of error behaviour and temporal dynamics, thereby minimizing redundancy among structurally similar approaches (e.g., multiple forest variants) and ensuring complementary inductive biases for subsequent ensemble construction (Table S1). This diversity was considered essential for constructing a robust weighted ensemble capable of adapting across heterogeneous climatic regimes. Six representative models were selected for ensemble development based on their out-of-sample predictive performance and complementary methodological diversity (Table 1): Regularized Linear Regression (RLR) implemented via elastic-net penalization, Extreme Gradient Boosting trees (XGB), Neural Network Autoregression (NNETAR), Random Forest regression (RF), Support Vector Regression with a radial basis function kernel (SVR), and a Prophet-XGBoost hybrid. In the latter, the Prophet model captures trend and seasonal components, while remaining residual structure ("errors") is subsequently modeled using XGBoost to account for

Table 1
Methodological characteristics of models included in the final ensemble. Overview of forecasting models retained after initial screening of 61 candidate algorithms. The table lists the conceptual model name used in the manuscript, its R implementation, the corresponding methodological category, and the primary modeling framework.

Selected Model	R Implementation	Methodological Family	Rationale for Inclusion in Ensemble
Regularized Linear Regression (RLR)	glmnet	Penalized linear regression	Provides stable shrinkage-based estimates and interpretable linear effects; robust under multicollinearity.
Extreme Gradient Boosting (XGB)	xgboost	Boosted decision trees	Captures nonlinear interactions and higher-order feature effects via sequential boosting.
Random Forest (RF)	ranger	Bagged decision trees	Reduces variance through bootstrap aggregation; strong performance in nonlinear, noisy settings.
Support Vector Regression (SVR)	kernlab	Kernel-based learning	Models nonlinear relationships using RBF kernel in high-dimensional feature space.
Neural Network Autoregression (NNETAR)	forecast:nnetar	Autoregressive neural network	Captures nonlinear temporal dependence and lag-based memory effects.
Prophet–XGBoost Hybrid	prophet + xgboost	Hybrid seasonal–machine learning	Decomposes trend/seasonality via Prophet and models residual nonlinear structure using boosting.
ARIMA (benchmark)	forecast:auto.arima	Classical time-series model	Provides parametric autoregressive baseline capturing serial correlation structure.

nonlinear effects not explained by the baseline model. The additional inclusion of automatically selected AutoRegressive Integrated Moving Average (ARIMA) family ensured a classical time-series benchmark to quantify the contribution of temporal autocorrelation relative to meteorological predictors. Where applicable, regression models with ARIMA errors were also employed, in which serial dependence in the residuals is explicitly modeled after accounting for exogenous predictors. ARIMA orders were determined automatically using an information-criterion-based selection procedure (minimizing AICc), allowing the autoregressive and moving-average structure to adapt to site-specific temporal dependence patterns. Consequently, optimal ARIMA orders differed across cities, reflecting differences in their underlying pollen time-series dynamics. The NNETAR(p, P, k)[m] hyperparameters (p the number of non-seasonal lagged observations, P the number of seasonal lags, k the number of neurons in the hidden layer, and m the seasonal period) were selected automatically using a data-driven procedure, whereby the lag structure was inferred from the time-series properties and the network size was optimized via internal resampling on the training data. This allows the model to adapt flexibly to site-specific temporal dynamics.

All predictors in $x_{s,t}$ were supplied to each base model, except where model-specific workflows required removal of the raw date variable. Preprocessing steps, such as dummy encoding of categorical predictors and normalization of continuous variables, were embedded within each workflow to ensure consistent application across algorithms (Fig. 2).

All data processing, modelling, and statistical analyses were conducted in R (R Core Team, 2022; version 4.4.1), using established packages for model implementation (e.g., glmnet, xgboost, forecast, nnetar, ranger, kernlab, prophet and auto.arima).

2.4. Ensemble construction, evaluation, and outputs

To combine the strengths of the seven selected individual sub-models, ensemble forecasts were generated using three aggregation strategies: a simple mean, a simple median, and a weighted ensemble assigning higher weights to better performing models. All three ensembles were based on the same set of predictors and preprocessing pipeline, differing only in the aggregation method. Among these, the weighted approach provided the best overall performance across most evaluation metrics and was therefore adopted for the final forecasts. Each model's weight ω_i was derived as the inverse of its RMSE, normalized such that the sum of weights equals one:

$$\omega_i = \frac{RMSE_i^{-1}}{\sum_{j=1}^{M} RMSE_j^{-1}}$$

where M is the total number of sub-models in the ensemble. The ensemble forecast at time t was then computed as the weighted sum:

$$\hat{y}_{ens,t} = \sum_{i=1}^{M} \omega_i \hat{y}_{i,t}$$

where $\hat{y}_{i,t}$ denote the forecast from model i at time t.

This weighting structure prioritises models with superior out-of-sample accuracy, while ensuring that all contributions remain well-defined and interpretable.

Weights were calculated independently for each monitoring site and device, using RMSE values obtained from the rolling-origin validation folds. This site-specific calibration allows the ensemble to adapt to local data characteristics (e.g., climatology, pollen season duration) while maintaining the same predictor structure and preprocessing pipeline across all component models.

Overfitting was mitigated by rolling-origin resampling, which preserves the temporal structure of the data and provides repeated out-of-sample validation windows. In addition, many of the component algorithms incorporate built-in regularization (e.g., L1/L2 penalties in RLR, shrinkage and early stopping in XGB, dropout-like regularization in NNETAR), further reducing the risk of overfitting to the training period.

Numerical forecasts of daily pollen concentrations were subsequently classified into categorical exposure levels using five predefined classes: No pollen [$\leq$3 pollen grains/m^3], Low [$>3 \leq 10$], Medium [$>10 \leq 30$], High [$>30 \leq 60$], and Very High [>60]. This classification scheme was applied uniformly across study sites to enable standardised comparison of model performance in predicting categorical pollen risk results across regions. The exposure thresholds were derived from published clinical and aerobiological studies (Buters et al., 2015; Cassagne et al., 2007; D'Amato et al., 2007; de Weger et al., 2013; Erbas et al., 2007) and correspond to commonly adopted operational *low--moderate-high-very high* categories used by European aerobiological monitoring networks, including the Spanish Aerobiological Network (REA, in Spanish [www.uco.es/rea]) (Galán et al., 2007).

All forecasts, forecasting power metrics, and model weights were compiled into structured tables to support comparative evaluation.

To quantify the contribution of individual predictors to the model performance, we calculated SHAP (SHapley Additive exPlanations) values for the main meteorological variables included in the forecasting

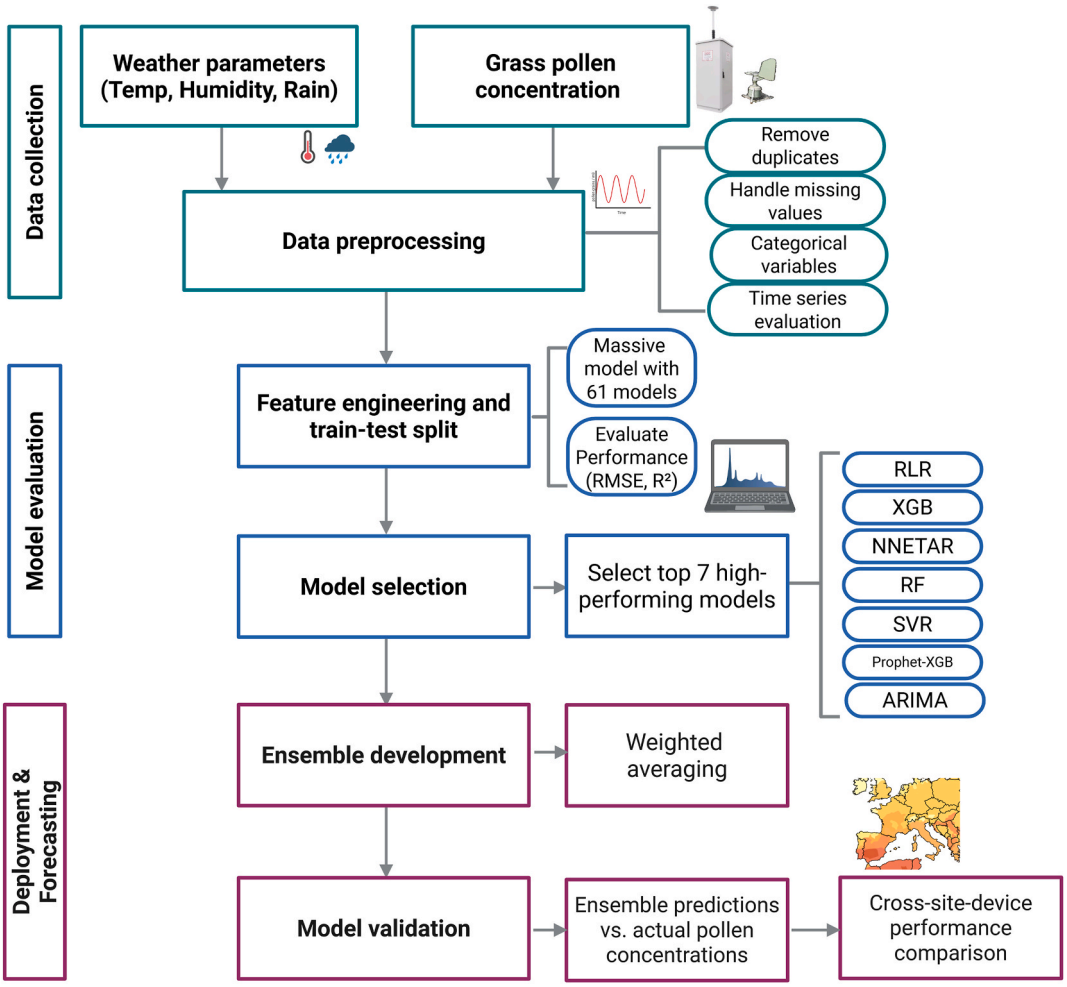

Fig. 2. Workflow for the development of the Poaceae pollen ensemble-forecasting model. Flowchart diagram illustrating the step-by-step process of developing the ensemble prediction model for Poaceae airborne pollen concentrations across the three study sites (Created in https://BioRender.com).

framework. SHAP provides a unified measure of feature importance by quantifying the marginal contribution of each predictor to the model's output, averaged across all possible predictor combinations, allowing the identification of the most relevant environmental drivers of airborne pollen concentrations. SHAP values were derived from the ensemble models and summarized across sites and devices to highlight consistent patterns in predictor relevance.

2.5. Sensitivity analysis

To evaluate the robustness of the forecasting framework, we conducted a multi-component sensitivity analysis examining.

i. **alternative ARIMA specifications.** Because automatic model selection can yield different parameterizations across stations, we evaluated three alternative ARIMA formulations: (a) an automatically selected ARIMA model (full parameter search with exogenous regressors), (b) a fixed ARIMA(4,0,0) model, and (c) a seasonal SARIMA(4,0,0)(1,0,0)[7] model incorporating weekly seasonality. In this notation, the first triplet (p,d,q) denotes the order of the non-seasonal autoregressive (p), differencing (d), and moving-average (q) components, respectively, while the second triplet specifies the corresponding seasonal components. The seasonal period [7] reflects weekly recurrence in daily data. All ARIMA variants were fitted using the same exogenous predictors as in the main analysis, and their predictive performance was compared using the independent test period to quantify sensitivity to model parameterization.

ii. **changes to ensemble composition.** To assess whether the ensemble performance was affected by the choice of constituent sub-models, we compared three configurations *per* station: Full ensemble, all available models weighted by inverse RMSE; reduced ensemble (no weak families), excluding RF, SVR, and Prophet-XGB, which were consistently weaker across stations; and Top-4 ensemble, containing only the four best-performing single models (ranked by test RMSE). For each configuration, we recomputed ensemble predictions on the held-out test set and quantified differences in RMSE and R^2 relative to the full ensemble. This analysis evaluates stability under alternative model-selection strategies.

iii. **different lag structures for pollen predictors.** Because airborne pollen concentration exhibits strong autocorrelation, we evaluated how different temporal lags affected predictive performance. For each station, three lag structures were examined: *lag-1*: pollen on the previous day; *lag-2*: pollen two days earlier; and *lag-3*: pollen three days earlier. For each lag setting, we constructed: $P_{s,t\text{-}lag}$ and a smoothed version: $(\overline{P}^{(3)}_{s,t-lag})$. We also defined a categorical variable ("PollenYN") indicating whether pollen had been observed recently, based exclusively on lagged values to prevent target leakage. All candidate models (RLR, XGB, NNETAR, RF, SVR, Prophet-XGB, and ARIMA-family with exogenous regressors) were retrained separately for each lag

configuration, and an inverse-RMSE weighted ensemble was constructed for each station × lag combination. This allowed us to quantify how assumptions about short-term temporal dependence influence both, single-model and ensemble accuracy.

iv. **device-level alignment at same site.** At the Augsburg location, where both a Hirst-type volumetric trap and an automatic PoMo sampler were available, we performed a device-level comparison to assess whether forecast accuracy depended on the measurement device. For each device, we computed the ensemble predictions on the independent test set and calculated daily forecast errors. Agreement between PoMo- and Hirst-based ensemble forecasts was evaluated using paired daily predictions. Forecast differences were summarized using the mean, standard deviation, and the 95% empirical intervals obtained via bootstrapping. Rank-based agreement between the two forecast series was assessed using Spearman's correlation, avoiding distributional assumptions. Additionally, we produced a Bland-Altman-style plot to visualise systematic bias and the spread of differences across the prediction range.

Each sensitivity component was evaluated under exactly the same rolling time-series validation framework used in the main analysis and the performance was assessed exclusively on out-of-sample data using RMSE, MAE, MASE and R^2 to ensure full comparability of results.

3. Results

3.1. Descriptive weather and pollen seasons overview

Grass pollen concentrations showed marked seasonal variation across the three study sites, with notable differences in timing, duration, and intensity between Augsburg, Córdoba, and Thessaloniki. In Augsburg, the season typically began in mid-May and extended into mid-August, lasting an average of 110 days. By contrast, the Córdoba season started earlier (mid-March) and was longer, extending into early August with a median length of 157 days. In Thessaloniki, the season started later than in Córdoba and earlier than in Augsburg (in early April), but it was much longer, often extending into early October. This resulted in a median season length of 200 days, the longest among the sites.

The timing of peak concentrations also differed: in Augsburg, the median peak occurred around the beginning of June (day of year, doy = 157). In Córdoba and Thessaloniki, the peaks occurred earlier, in mid-May, at 133 doy and 139 doy, respectively (see Table 2, Table S2 and Fig. S1).

Pollen intensity was moderate in Augsburg, with the Hirst trap recording a mean Seasonal Pollen Integral (SPIn) of approximately 1000 pollen grains/m^3. Córdoba exhibited substantially higher pollen loads, with mean SPIn values around 6000 pollen grains/m^3 and generally more intense peak periods. Thessaloniki showed the lowest overall pollen intensity among the sites, with a mean SPIn of roughly 550 pollen grains/m^3 despite its comparatively long pollen season (Fig. S1).

Across the study period (2018-2024), Augsburg exhibited relatively cool and humid pollen-season conditions, with mean temperatures around 16-18 °C, high relative humidity (≈70-77 %), and frequent rainfall, typically exceeding 200-500 mm *per* season. Córdoba showed the warmest and driest climate, with mean pollen-season temperatures of 20-23 °C, low humidity (≈48-55 %), and very limited precipitation, accompanied by long dry spells often lasting several weeks. Thessaloniki presented intermediate conditions, with mean seasonal temperatures of 21-23 °C, moderate humidity (≈56-66 %), and variable rainfall ranging from very dry years to seasons with over 160 mm of precipitation (Fig. S2, Table S3).

Table 2
Summary of pollen season parameters for Augsburg, Cordoba and Thessaloniki (2018–2024). Values represent medians across years for start and end dates, season length (days), day of year (doy) of concentration peak, and mean Seasonal Pollen Integral (SPIn).

station	Median start date	Median end date	Median length (days)	Median peak date (doy)	Mean SPIn
Augsburg	15/05	19/08	110	157	995.5
Córdoba	17/03	10/08	157	133	5994.1
Thessaloniki	06/04	05/10	200	139	553.7

3.2. Performance of sub-models across locations

The evaluation of the seven forecasting algorithms revealed distinct patterns in predictive performance across sites and devices. Predictive performance, evaluated using RMSE, MAE, MASE, and R^2, showed that no single model dominated universally, but some consistent trends emerged. Across all evaluation metrics, predictive performance was consistently best in Thessaloniki, intermediate in Augsburg, and weakest in Córdoba (Fig. 3, Table S4).

In Augsburg the best-performing submodel was ARIMA(4,0,0), which achieved RMSE = 5.47 pollen grains/m^3 (R^2 = 0.671, MAE = 1.83 pollen grains/m^3, MASE = 0.56) and received the highest ensemble weight (0.18). RLR and XGB ranked next (RMSE 5.58-5.79 pollen grains/m^3; R^2 0.65), forming a close secondary tier (Fig. 3, Table S4). Prophet-XGB, SVR, and RF produced moderate errors (RMSE 6.6-7.0 pollen grains/m^3), whereas NNETAR showed substantial underperformance (RMSE ≈ 25.6 pollen grains/m^3; R^2 = 0.09), contributing minimally to the ensemble weight (0.04). The strong performance of ARIMA extended across all error metrics and was the only model with consistently low MAE and MASE at this site.

Córdoba showed the highest overall errors, reflecting more intense and variable pollen seasons. Here, the NNETAR(4,1,10)[7] model performed best, with RMSE = 42.04 pollen grains/m^3, R^2 = 0.688, and the highest ensemble weight (0.17). RLR (RMSE = 46.80 pollen grains/m^3; R^2 = 0.587) and ARIMA(5,0,0) (RMSE = 46.95 pollen grains/m^3; R^2 = 0.583) formed an intermediate-performing cluster. Tree-based (XGB, RF) and kernel-based (SVR) models generated larger errors (RMSE ≈ 53-61 grains/m^3) and lower R^2 (0.30-0.46). Most Córdoba models had MASE >1, suggesting higher normalized error relative to Augsburg and Thessaloniki (Fig. 3, Table S4). Although R^2 values in Córdoba were comparable to Augsburg for some models, absolute and normalized error metrics (RMSE, MAE, MASE) were substantially higher, confirming lower overall predictive accuracy.

Thessaloniki showed error levels comparable to Augsburg and lower than Córdoba. NNETAR(4,1,10)[7] was the best-performing model, reaching RMSE = 2.65 pollen grains/m^3 (R^2 = 0.908, MASE = 0.41), and the largest ensemble weight (0.20). RLR and ARIMA(5,0,0) followed closely (RMSE ≈ 3.22-3.25 pollen grains/m^3; R^2 ≈ 0.80), and XGB delivered intermediate performance (Fig. 3, Table S4). RF, SVR, and Prophet-XGB had RMSE between 4.18 and 4.36 pollen grains/m^3, with R^2 ≈ 0.69-0.77. As in Augsburg, tree-based and kernel methods ranked below ARIMA and linear models.

Importantly, the ARIMA model specification differed among stations (e.g., ARIMA(4,0,0) in Augsburg, ARIMA(5,0,0) in Córdoba and Thessaloniki). This variation reflects the automatic parameter selection of the auto-ARIMA optimisation procedure, which adapts model structure to local temporal dynamics and information-criterion minimisation. To examine whether these station-specific differences in ARIMA structure had any measurable impact on predictive performance, we conducted a dedicated sensitivity analysis comparing fixed ARIMA formulations (Section 3.4.1).

The heterogeneity in sub-model performance across stations, with different algorithms achieving the best accuracy in each location provided the rationale for using a weighted ensemble-forecasting framework.

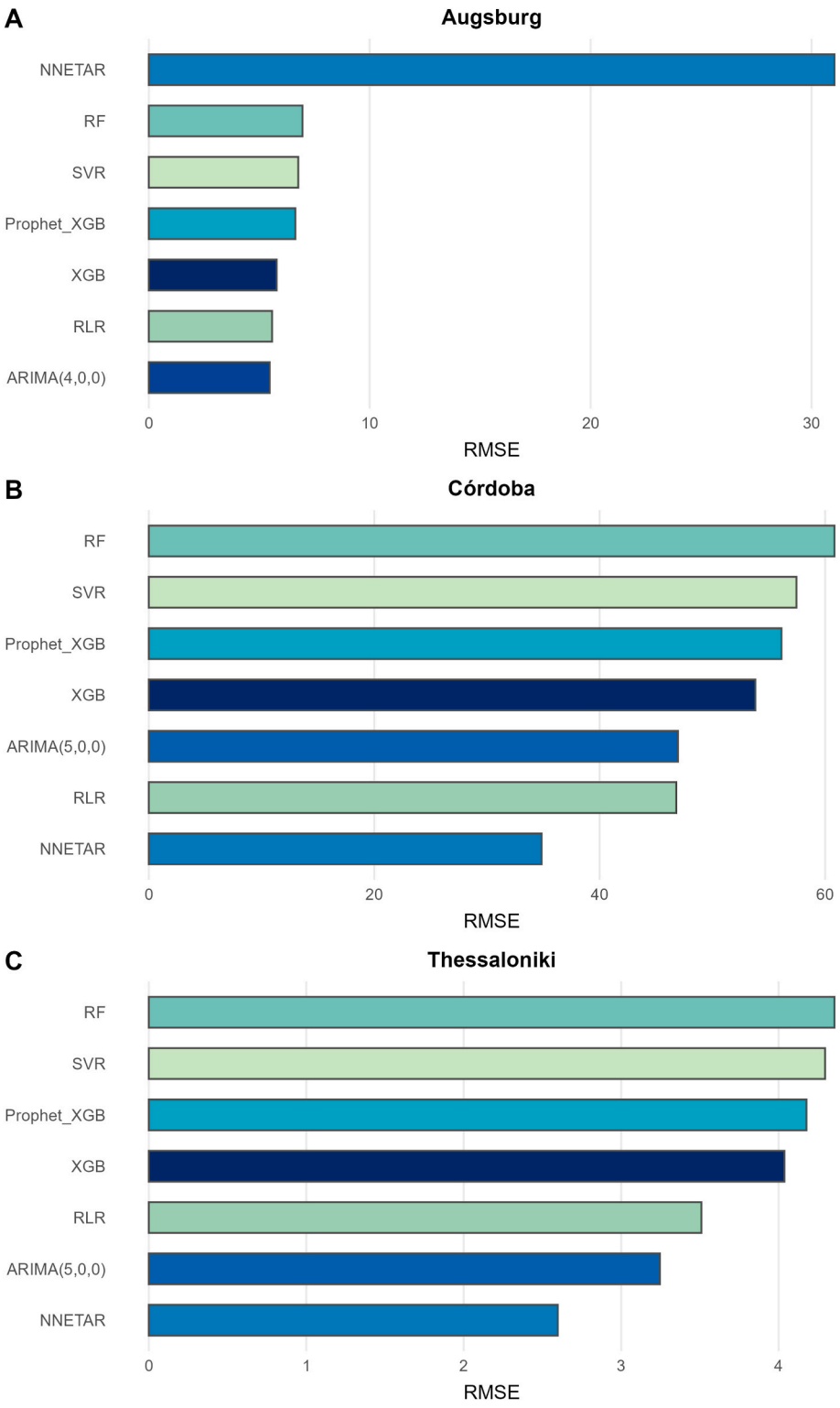

Fig. 3. Performance of individual forecasting models across the three stations. Sub-model performance (RMSE) for (A) Augsburg, (B) Córdoba, and (C) Thessaloniki with pollen concentration obtaining by the Hirst-type device. Colours are consistent across panels for each algorithm; lower RMSE indicates better predictive accuracy.

3.3. Ensemble performance and classification

During the independent validation year 2024, the inverse-RMSE weighted ensemble, combining all seven submodels, produced stable and competitive forecasts across the three stations. Ensemble performance varied according to local pollen dynamics, with error magnitudes reflecting site-specific 2024 season intensity.

In Augsburg, the ensemble attained RMSE = 5.74 pollen grains/m^3, R^2 = 0.662, MAE = 2.32 pollen grains/m^3, MASE = 0.71 (Fig. 4, Table S5). These results show that the ensemble performed closely to the best individual submodels (e.g., ARIMA and RLR), providing a balanced representation of the station's relatively moderate pollen

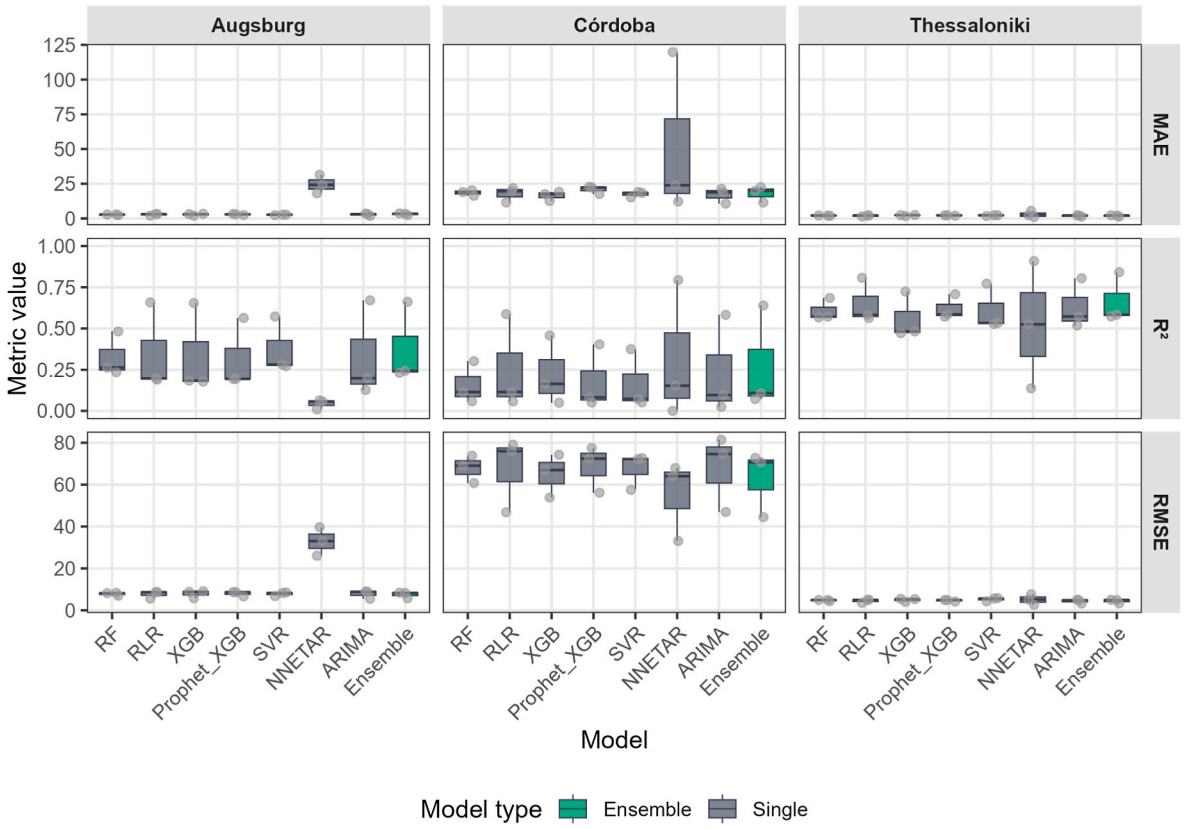

Fig. 4. Model accuracy to lag structure across stations. Boxplots and jittered points show the distribution of forecasting accuracy (MAE, R^2, RMSE) for individual sub-models (grey) and the weighted ensemble (green) across three stations (Augsburg, Córdoba, Thessaloniki) and three lag configurations (1-3 days). Columns represent the combination of station and lag, and rows correspond to each metric. For R^2, values are constrained to the [0, 1] range for comparability across panels.

concentrations. The ensemble time series closely followed observed fluctuations, particularly during peak episodes in May-June, and reproduced the low-level background season accurately (Fig. 5).

The ensemble exhibited higher errors in Córdoba, consistent with the more intense and abrupt pollen dynamics at this site. The model achieved RMSE = 45.61 pollen grains/m^3, R^2 = 0.617, MAE = 12.38 pollen grains/m^3, MASE = 0.77 (Fig. 4, Table S5). Despite higher absolute values, ensemble performance represented a compromise between strong nonlinear models (NNETAR) and linear components (RLR, ARIMA), providing stable predictions across the extended high-season interval. The confusion matrix indicated good classification of low and medium categories, with misclassifications concentrated around the highest pollen levels (Fig. 5).

In Thessaloniki, the ensemble achieved the best performance among the three stations RMSE = 3.34 pollen grains/m^3, R^2 = 0.842, MAE = 1.37 pollen grains/m^3, MASE = 0.58 (Fig. 4, Table S5). Ensemble forecasts reproduced the principal seasonal peak with high fidelity and maintained accuracy throughout lower-concentration periods. Classification accuracy for categorical pollen levels was also high, with most observations in the "No Pollen" and "Low" categories correctly identified (Fig. 5).

Across all stations, the ensemble consistently ranked within or very close to the top-performing individual models (Fig. 5), providing smoother and more stable predictions than any single algorithm. The ensemble effectively integrated the strengths of site-dependent sub-models, including ARIMA in Augsburg, NNETAR in Córdoba and Thessaloniki, and RLR at all sites, resulting in robust performance across diverse climatic and phenological contexts.

The ensemble provided accurate reconstructions of the 2024 grass pollen season, though performance varied in magnitude (Fig. 5, Table S5). At Augsburg, the ensemble achieved a moderate model fit (R^2 = 0.66), capturing the timing and shape of the main pollen season but showing partial underestimation of the highest peaks. Despite this, categorical performance remained strong: most "no pollen" and "low" days were classified correctly, and the model reliably separated low from medium and high exposure levels in the confusion matrix. At Córdoba, the ensemble reproduced the overall seasonal pattern (R^2 = 0.62) but showed reduced accuracy for very high-intensity days, with several extreme peaks underpredicted. Even so, categorical forecasts demonstrated clear discrimination between medium, high, and very high levels, indicating that the model maintained useful level-based agreement despite larger numerical errors. At Thessaloniki, ensemble performance was markedly higher (R^2 = 0.84), with close alignment between observed and predicted concentrations throughout the season. The model accurately reconstructed both the main seasonal peak and the surrounding lower-concentration periods. Categorical performance was also high, with strong agreement across the full range of pollen levels. The confusion matrix revealed high reliability in identifying no/low pollen days as well as medium-level exposures (Fig. 5).

SHAP-based feature importance analysis showed that the strongest drivers of forecasted pollen levels were the pollen-related predictors, particularly 1-day lag and the 3-day moving average of past grass pollen concentration, as well as date-derived seasonal signatures. These variables consistently dominated across all stations, reflecting the strong autoregressive and seasonal structure of grass pollen dynamics (Fig. S3). Meteorological predictors contributed additional explanatory power but ranked below the pollen-derived features. In Augsburg and Córdoba, temperature was the most influential weather variable, followed by relative humidity, while precipitation showed a smaller but still significant effect. In Thessaloniki, meteorological influences were noticeably weaker overall: humidity emerged as the dominant weather predictor, followed by precipitation, and temperature contributing least (Fig. 6).

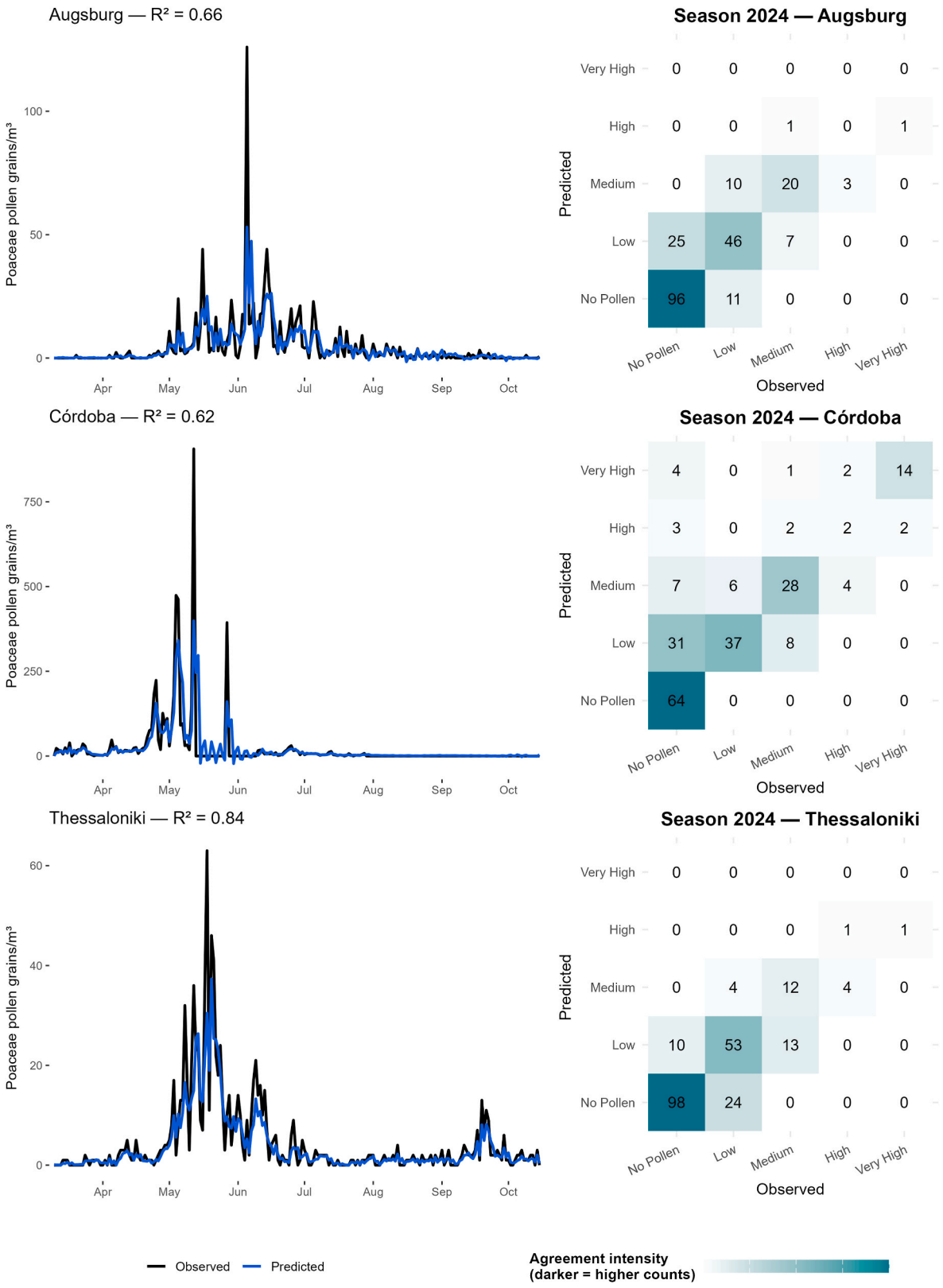

Fig. 5. Observed versus ensemble-predicted pollen concentrations and categorical accuracy. Ensemble forecasts of daily Poaceae pollen concentrations during the 2024 grass pollen season. Left panels show observed (black) versus predicted (blue) pollen counts for each site (Augsburg, Córdoba, Thessaloniki), with the coefficient of determination (R^2) reported as an overall measure of fit. Right panels present corresponding confusion matrices after converting numerical forecasts into categorical pollen risk levels (No Pollen, Low, Medium, High, and Very High). Darker cells indicate higher counts, and alignment along the diagonal reflects agreement between observed and predicted categories.

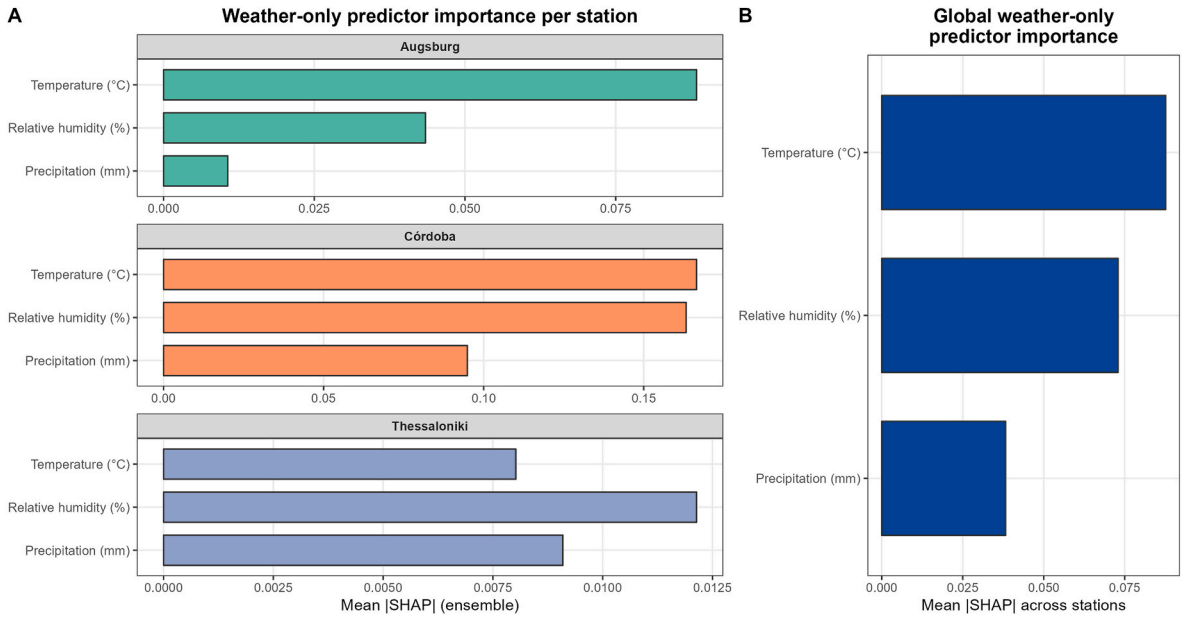

Fig. 6. SHAP-based weather predictor importance at station level and globally. SHAP (SHapley Additive exPlanations) weather-only feature importance for predicting daily Poaceae pollen concentrations using the ensemble forecast in (A) Station-specific mean absolute SHAP values for temperature, relative humidity, and precipitation across Augsburg, Córdoba, and Thessaloniki, showing how the relative contribution of each meteorological driver varies by climate and local pollen dynamics. (B) Global weather-only importance aggregated across the three stations. Bars represent the mean absolute SHAP value of each environmental predictor, quantifying its average contribution to model output. Larger bar lengths indicate higher predictive importance.

3.4. Ensemble diversity and sensitivity analysis

3.4.1. Model-specific sensitivity: ARIMA comparison

A comparison of the three ARIMA variants tested (a non-seasonal ARIMA(p,0,q), the alternative selected by auto-ARIMA, and a seasonal SARIMA extension with weekly seasonality) showed that the forecasting metrics were nearly identical across all sites (Fig. S4). This indicates that the performance of the model was largely insensitive to moderate changes in the autoregressive order or the inclusion of a simple seasonal term. In Augsburg, the tested configurations (ARIMA(4,0,0) and SARIMA(4,0,0)(1,0,0)[7]) produced virtually indistinguishable results (RMSE = 5.47-5.47 pollen grains/m^3; R^2 = 0.6707-0.6709) (Fig. S4). Córdoba showed the same pattern: ARIMA(5,0,0), ARIMA(4,0,0), and SARIMA(4,0,0)(1,0,0)[7] all produced an RMSE ≈ 46.94 pollen grains/m^3 and R^2 ≈ 0.583, with differences below the third decimal place. Thessaloniki likewise displayed highly stable behaviour, with all three ARIMA structures achieving an RMSE ≈3.246 pollen grains/m^3 and R^2 ≈ 0.8044. These results confirm that ARIMA-based learners are robust to reasonable changes in model structure and that the auto-ARIMA selections used in the main ensemble analysis are consistent with near-optimal local configurations.

3.4.2. Alternative ensemble configuration performance

The evaluation of alternative ensemble configurations revealed that the inclusion or exclusion of weaker learners only moderately affected forecast power, although the magnitude and direction of this effect varied across stations. At the Augsburg site, the full 7-model ensemble achieved an RMSE of 5.74 pollen grains/m^3 (R^2 = 0.662), but excluding the weakest model families (i.e., retaining RLR, XGB, NNETAR, and ARIMA-family) slightly improved explanatory power (R^2 = 0.668) while reducing RMSE to 5.50 pollen grains/m^3. The Top-4 ensemble, composed of the four sub-models with the best local performance, provided the lowest MAE (1.95 pollen grains/m^3) and MASE (0.60). However, there was only a marginal improvement in R^2 (0.676), indicating that the performance benefits were small and might have largely been driven by central-tendency errors rather than by systematic improvements in the fit (Fig. S5).

In Córdoba, where single-model performance variability was highest, the ensemble restructuring produced more noticeable differences. Removing weaker families reduced RMSE from 45.61 (full ensemble) to 40.86 pollen grains/m^3, an 11% improvement, and increased R^2 from 0.617 to 0.687. The Top-4 ensemble performed similarly (RMSE 41.64 pollen grains/m^3, R^2 = 0.675), suggesting that most of the benefits arose from omitting systematically underperforming learners rather than from restricting the ensemble to the very best models. The ensemble models in Córdoba are particularly sensitive to the inclusion of low-performing algorithms, which is consistent with the greater variation observed in the accuracies of individual models at this site.

In Thessaloniki, all ensemble structures performed strongly, but pruning consistently improved the accuracy. The full ensemble achieved an RMSE = 3.34 pollen grains/m^3 and R^2 = 0.842 but removing the weak model families resulted in an even better performance (RMSE = 2.96 pollen grains/m^3, R^2 = 0.870). Although the Top-4 ensemble showed slightly higher RMSE (3.11), it still improved R^2 relative to the full ensemble (Fig. S5). Thessaloniki's strongest base-learners (NNETAR, RLR, ARIMA) performed consistently well, so ensembles relying more heavily on them benefit from reduced noise.

Overall, the full seven-model ensemble delivered consistently competitive performance across all stations. Ensembles excluding weaker model families produced modest improvements at some stations, particularly where sub-model performance was more heterogeneous. However, these gains were not consistent across sites or metrics and were generally small.

3.4.3. Lag sensitivity

Across stations, increasing the prediction lag (1-3 days) produced substantial performance degradation for several individual models, particularly neural networks, Prophet-based models, and ARIMA variants (Fig. S6, Table S6). In both Augsburg and Córdoba, several learners showed sharp increases in RMSE and decreases in R^2 at lag 2 and lag 3, indicating that the impact of lag is strongly model- and site-dependent. Thessaloniki was the most stable site, with moderate performance losses in most learners.

In contrast, the weighted ensemble remained comparatively robust

across lag settings. Increasing the prediction lag (1-3 days) produced substantial performance degradation for several individual models, particularly neural networks, Prophet-based models, and ARIMA variants. Thessaloniki was the most stable site, with moderate performance losses in most learners. The weighted ensemble RMSE increased with lag (Augsburg: 5.75 → 8.41 pollen grains/m^3; Córdoba: 44.46 → 72.80 pollen grains/m^3; Thessaloniki: 3.33 → 4.99 pollen grains/m^3), yet the ensemble consistently ranked among the best-performing models at all stations and lags.

Importantly, the sensitivity analysis confirmed that explicit lagged predictors beyond 1 day (and the 3-day moving average) did not improve performance for any station, supporting the use of a compact set of lagged pollen features in the final model. Although ARIMA error structures occasionally selected higher autoregressive orders (e.g., AR = 4-5), these reflect residual autocorrelation dynamics rather than a need for longer lagged predictors in the regression component.

3.4.4. Device-based sensitivity: PoMo vs Hirst (Augsburg)

When each device was used as both training and evaluation target, both daily ensembles achieved good fits: the Hirst-based ensemble reached RMSE = 5.74 pollen grains/m^3 and $R^2 = 0.66$, while the PoMo-based ensemble trained on PoMo counts yielded RMSE = 8.20 pollen grains/m^3 and $R^2 = 0.89$. The comparison between both devices-based ensemble forecasts indicates strong overall agreement. Spearman rank correlation indicated a high association between the two forecast series ($\rho = 0.77$, $p < 0.001$), despite systematic differences in magnitude, while Pearson correlation was higher (r = 0.87), reflecting sensitivity to extreme values. Normality testing of forecast differences rejected Gaussianity (Shapiro-Wilk $p < 0.001$), supporting the use of rank-based metrics. PoMo-based forecasts were, on average, 9.16 pollen grains/m^3 higher than Hirst-based forecasts, with a wide empirical 95% interval of differences (−1.84 to 50.19) (Fig. S7). Importantly, this analysis does not assume either device as the true reference but instead quantifies how similarly the two systems behave when used to drive identical forecasting models.

4. Discussion

In this work, we aimed at creating a reliable and robust forecasting model of airborne pollen concentrations, with several real-life applications for environmental and public health sciences. Towards the prospect of a universal model to be applicable, or customisable, for a variety of distinct climatic areas, we used datasets from Augsburg/Germany, Córdoba/Spain and Thessaloniki/Greece. We originally hypothesised that such a theoretical framework might not be applicable to contrasting ecological locations. Our findings reported here demonstrate that pollen forecasting performance is indeed highly site-dependent, underscoring the limitations of relying on a single modelling approach across diverse environments. This site-dependence also reflects the strong ecological imprint of local topography, vegetation, phenology, and climate, all of which shape the temporal structure of airborne pollen. In this sense, pollen forecasting acts not only as a predictive tool but also as an indirect descriptor of ecosystem functioning across contrasting European ecoregions, as previously highlighted in ecological informatics approaches to vegetation-atmosphere interactions (Navares and Aznarte, 2020; Rojo et al., 2022). So, while one may initially conclude that generalised predictive models may not be feasible across different geographically and climatically sites, the truth seems to be that this is possible under customisable environmental parameters.

Different algorithms dominated across sites: in Augsburg, ARIMA provided the best predictive performance while NNETAR collapsed; in Córdoba and Thessaloniki, both NNETAR and ARIMA performed strongly, whereas tree-based models such as RF and XGB were clearly weaker. These differences reflect local variability in pollen dynamics, land-use patterns, meteorological forcing, and signal-to-noise ratios, which determine how well each algorithm captures temporal and environmental dependencies. This aligns with the wider literature showing that model skill often degrades when transferred across locations without local tuning (Suanno et al., 2021). Nevertheless, our results here point out the potential of using semi-generalised forecasting models for similar bioclimatic regions and with comparable predictive power. However, it is imperative to exercise caution and refine the models, particularly during the initial development phase, within the context of adaptive, location-sensitive forecasting systems. Importantly, the need for local tuning mirrors well-known ecological principles, as pollen production and phenology respond non-linearly to regional climate and land use. This reinforces the ecological relevance of forecasting models that can adapt to local biological and environmental drivers. The strong lag sensitivity observed in some learners further confirms that memory effects on pollen time series are process-dependent, whereas the ensemble architecture mitigates this instability by down-weighting models that deteriorate at longer horizons while retaining stabilizing components such as ARIMA and RLR.

The comparative results also show that no single model family consistently dominates. ARIMA (or SARIMA), commonly treated as a baseline (Hurtado et al., 2024; Muzalyova et al., 2021; Rodriguez-Rajo et al., 2006), was competitive at two sites and outperformed machine-learning models in some metrics. Rankings were broadly consistent across RMSE, MAE, R^2 and MASE, although the relative ordering of models varied slightly depending on the error measure considered. This reinforces the importance of evaluating forecasts with multiple complementary metrics, as increasingly recommended in time-series forecasting research (Makra et al., 2024).

Climatic contrasts at the three stations were directly reflected in both, pollen-season structure and model behaviour. Although Córdoba and Thessaloniki are both classified within the broader Mediterranean domain, their environmental and land-use conditions differ in ways that are highly relevant for grass pollen dynamics. Córdoba's hot, dry Mediterranean climate, shaped by its inland location within the Iberian Peninsula and strong continental influence, and combined with extensive open agricultural landscapes, produced early, sharp, and intense pollen peaks, favouring models capable of capturing nonlinear pulses (e.g., autoregressive and tree-based methods). Thessaloniki, by contrast, is moderated by its proximity to the Aegean Sea; this maritime influence results in more regulated temperatures, higher humidity, and regular rainfall, combined with more heterogeneous urban and peri-urban vegetation. These conditions generate extended flowering periods with smoother daily variability, under which season-structured models such as ARIMA and RLR performed particularly well. Augsburg's temperate to humid-continental setting, characterised by cooler temperatures, higher humidity, frequent rainfall, and a sparser distribution of grassland habitats, produced shorter and less intense seasons with greater stochastic noise; accordingly, regularized and error-correction models (e.g., regression with ARIMA errors) performed best. Importantly, differences in seasonal pollen integrals and peak intensities also imply differing levels of forecasting difficulty. Córdoba exhibited substantially higher SPIn values and sharper peak concentrations compared with Augsburg and Thessaloniki, resulting in greater variance and larger absolute error magnitudes. In contrast, the lower seasonal integrals in Augsburg and Thessaloniki led to a higher proportion of low or zero-concentration days, introducing class imbalance in categorical evaluations. These structural differences in pollen climatology partly explain the site-specific variation observed in both continuous and categorical performance metrics. Across these contrasting regimes, our weighted ensembles systematically produced narrower error distributions, more stable R^2 values across lags, and greater robustness compared with single models. In practice, the ensemble protected against the collapse of individual algorithms, such as NNETAR in Augsburg or tree-based learners in Córdoba, by leveraging the complementary strengths of other models. The limited and inconsistent gains observed when pruning weaker model families further confirm that the inverse-RMSE weighting scheme already suppresses the influence of underperforming learners,

allowing the full weighted ensemble to remain a robust default option across contrasting environments. The lag analysis shows that the accuracy of forecasts is not consistent for individual learners across different lag settings. However, the ensemble architecture mitigates this variability largely maintaining competitive performance even under more challenging forecasting horizons. This kind of ensemble stability has been documented in other aeroallergen forecasting contexts, such as *Ambrosia* pollen, where combining deep neural networks, random forests and boosting improved performance and resilience (Zewdie et al., 2019). These patterns also underline how climate-driven shifts in temperature and precipitation regimes may alter pollen season structure in the future, making adaptable forecasting frameworks increasingly valuable under ongoing climate change. From an ecological viewpoint, this robustness is critical, as future shifts in land use and climate are expected to reorganise grass phenology, pollen productivity, and dispersal connectivity across landscapes. In this context, pollen time series act as dynamic bioaerosol signals reflecting vegetation phenology, ecosystem productivity, and the biological fingerprints of climate change and land-use transitions. At the same time, our results also emphasize that statistical generalisation is inherently constrained by ecological stationarity: pollen-climate relationships derived at one site or ecoregion cannot be assumed to transfer unchanged across landscapes with different vegetation composition, management intensity, or bioclimatic forcing, even worse if one considers such process at both a macro-ecology and micro-ecology scale.

An additional advantage of ensemble modelling is its capacity to provide insight into predictor importance and model interpretability. By applying SHAP, we identified lagged pollen concentrations and temperature as the strongest drivers across sites, followed by humidity and precipitation categories. Similar findings have been reported for olive and birch pollen, where temperature and rainfall played dominant roles in season dynamics (García-Mozo et al., 2014; Grewling et al., 2012; Rojo et al., 2021). Interpretable models not only improve scientific understanding but also enhance trust among health professionals and end-users; for example, Wang et al. (2024) and Holzmann et al. (2025) used SHAP-based feature ranking in allergic rhinitis prediction to increase the clinical usability of their model. Improved interpretability is especially relevant in clinical contexts, where transparent links between environmental drivers and predicted symptom risk help bridge environmental modelling with patient-centered care. Wind was not included among predictors because our models used a harmonized daily meteorological input set across stations and instruments, while wind effects on pollen at the city scale are often better represented by wind direction/persistence and air-mass pathways rather than mean daily wind speed alone; such transport effects have been shown to strongly modulate pollen inflow, including for Thessaloniki (Damialis et al., 2005; Picornell et al., 2023). Additionally, several statistical forecasting studies have reported that mean daily wind speed may show weak or non-significant relationships with daily pollen concentrations in some settings, particularly when other covarying meteorological variables and temporal dependence are already accounted for (Muzalyova et al., 2021). In addition, convective storm situations can alter the relationship between measured whole-grain pollen counts and allergenic particle load through pollen rupture and the release of sub-pollen particles, a mechanism widely linked to thunderstorm-asthma episodes and not fully captured by daily-mean predictors (Emmerson et al., 2021; Hughes et al., 2020). Future extensions of the proposed framework could therefore integrate higher-temporal-resolution wind metrics, wind direction or trajectory-based indicators, and proxies for convective storm activity to better represent these processes.

Our results align with and extend recent advances in pollen forecasting. Astray et al. (2025) applied machine learning methods including Random Forest, Support Vector Machine (SVM), and neural networks to forecast *Parietaria* pollen in Spain, obtaining correlations of 0.71-0.86 and RMSE values of 5.5-7.7 pollen grains/m^3. Picornell et al. (2024) implemented LSTM-based deep learning on monthly integrals of *Olea* and Urticaceae, showing strong improvements over traditional statistical models. On a broader scale, recent studies have shown that hybrid machine-learning approaches can improve both spatial pollen prediction and short-term forecast robustness under heterogeneous environmental conditions, including regional mapping in Switzerland (Valipour Shokouhi et al., 2024a) and multi-city forecasting applications (Makra et al., 2024; Yin et al., 2026), revealing that ensemble/dl methods outperformed simpler models in many but not all locations. Goudarzi et al. (2022) demonstrated that ANNs can forecast total pollen concentrations in challenging climates like Ahvaz (Iran), contributing evidence that non-linear models remain viable in less temperate settings. Finally, Lo et al. (2021) developed a Random Forest model to predict daily pollen levels in South Spain, showing that machine learning can outperform simpler regressions in operational contexts. Together, these studies point to a field evolving from single statistical models toward hybrid, machine learning-driven strategies. Our results contribute by demonstrating that ensemble approaches offer the most consistent and reliable forecasts across heterogeneous sites.

From a practical standpoint, the implications are clear. Pollen forecasting systems intended for operational use must anticipate the heterogeneity of local pollen dynamics and avoid dependence on a single algorithm. The inverse error weighting strategy applied here is widely used in time-series forecasting to integrate heterogeneous model families with complementary strengths while naturally limiting the influence of underperforming learners (e.g., Liang et al., 2022). Weighted ensembles represent a promising solution, as they integrate the autoregressive strengths of ARIMA, the nonlinear adaptability of NNETAR, the interaction learning capacity of tree-based models, and the stability of regularized regressions. In addition, explainable ensembles provide day-level attribution of which predictors drive forecasted risk (e.g., lagged pollen vs. meteorological factors), offering public health authorities actionable insights for targeted communication. Such interpretability improves end-user compliance, as forecast warnings are easier to trust when linked to understandable environmental drivers (Zhu et al., 2024). Beyond overall accuracy, the practical value of pollen forecasts also depends on their categorical reliability. Our results showed that severe misclassifications, such as predicting high risk on no-pollen days or failing to detect very high events, were uncommon across sites, although some overprediction in the "high" category was observed in Córdoba, which represents a structurally more challenging setting due to its markedly higher pollen intensities and sharper peak dynamics. Nevertheless, most predictions aligned along the diagonal in the confusion matrices, demonstrating robust identification of the general exposure levels most relevant for daily decision-making.

Off note, the public health relevance of robust pollen forecasting is underscored by recent evidence showing that personalised, data-driven pollen information can meaningfully reduce symptom burden in allergic individuals. In a recent randomised trial, Holzmann et al. (2025) demonstrated that an app providing a local grass pollen forecast, based on the forecasting framework developed in this study led to significant clinical benefits, including lower symptom severity, increased medication use, and improved overall disease control. These findings show that improvements in predictive performance do not just represent methodological advances but also translate directly into measurable health outcomes. Importantly, the pollen risk categories adopted in this study align with clinically and ecologically validated exposure thresholds for grass pollen (≈2-10 grains/m^3 for symptom onset; ≥30-50 grains/m^3 for high-risk days), reinforcing that the model outputs are biologically and medically meaningful. Given that climate change is expected to intensify pollen seasons and increase exposure levels, the clinical value of accurate, real-time forecasting is likely to become even more important in the coming decades.

Our comparison of forecasts generated using PoMo versus Hirst observations highlights the practical importance of device-specific performance for operational pollen prediction. On average, PoMo-based forecasts were higher than Hirst-based forecasts, with significantly

better performance. This positive bias is consistent with the higher pollen concentrations recorded by PoMo samplers. These operate at a higher volumetric flow rate than standard Hirst traps (10 L/min) and therefore might capture larger airborne pollen loads (Plaza et al., 2022). Although Hirst data remain the reference standard for aerobiological monitoring, PoMo-based ensembles performed markedly better in our lag-1 configuration, precisely the setting required for real-time forecasting. Because PoMo delivers automated, high-temporal-resolution data without the processing delay inherent to Hirst microscopy, the superior lag-1 performance achieved with PoMo has direct operational relevance. In practice, this means that PoMo-driven forecasts can provide same-day or next-day airborne pollen information with higher accuracy, enabling more effective and timely warnings. Thus, for real-time applications, automated samplers are not only logistically advantageous but can also yield stronger predictive models when appropriately integrated into an ensemble framework. At the same time, while PoMo-based results demonstrate the potential for accurate same-day and next-day pollen forecasts, we acknowledge several important limitations relevant for operational early-warning systems. Extending lead times beyond one to two days remains challenging, as predictive skill depends strongly on the availability and reliability of meteorological inputs. In the present study, the 2024 evaluation used observed weather data to assess intrinsic model performance under controlled conditions. In operational deployment, however, meteorological forecasts are required (as implemented in Holzmann et al. (2025), and overall predictive skill therefore reflects the combined uncertainty of both weather forecasts and the pollen model itself. Consequently, forecast accuracy may decrease with increasing lead time, particularly beyond short-term horizons. Forecasts are produced at monitoring locations and do not provide continuous pollen coverage across regions, limiting their direct applicability for regional exposure mapping. However, importantly, the proposed ensemble framework is not site-specific in structure: the same sub-models are applied across all locations, with site adaptivity achieved through performance-based weighting that reflects local pollen dynamics. Future developments could therefore focus on hybrid approaches that combine such adaptive ensemble learning with atmospheric dispersion or trajectory models and data assimilation techniques, enabling longer lead times and spatially resolved pollen forecasts while retaining the robustness of site-level predictions.

Together, our findings illustrate that reliable pollen forecasting requires models that adapt to local environmental complexity while remaining generalizable across regions. By combining methodological robustness, interpretability, and compatibility with automated monitoring systems, our ensemble framework represents a step toward scalable, real-time pollen intelligence. Future extensions could integrate land-use data, finer scale meteorological inputs, and cross-site transfer learning to further enhance performance. Ultimately, these advances strengthen the capacity of aerobiological forecasting directly to support public-health protection and climate-sensitive exposure management, while also offering complementary ecological indicator of shifting vegetation-atmosphere coupling across landscapes.

5. Conclusions

Our results show that while airborne pollen forecasting performance is expectedly site-dependent, there is a clear potential for transferable and scalable predictions once regional climatic, vegetation, and land use parameters are appropriately tuned. This variability highlights the risks of relying on any single algorithm for operational forecasts, and illustrates the clear advantage of ensemble models. The strength of the provided ensemble lies in its ability to deliver consistent and reliable predictions across heterogeneous bioaerosol environments, something no single algorithm achieved consistently. Whereas ARIMA, NNETAR, and tree-based models each showed strong performance at certain sites but failed at others, the ensemble successfully balanced these fluctuations by adaptively weighting models according to their strengths. This translated into narrower error distributions, more stable R^2 values across lags, and robust categorical risk predictions that align with the needs of health information providers and allergy sufferers. At the same time, limitations remain, including the site-level nature of the forecasts, the lack of explicit spatial representation, and constraints on extending lead times beyond one to two days using purely data-driven approaches. Towards resilient ecosystems, such ensemble forecasting approaches can provide in the future an operational tool for predicting climate-sensitive bioaerosols under increasingly unstable environmental conditions, especially in urban settings. If one additionally integrates real-time, automated biomonitoring, the resulting framework system is well suited to support new-era challenges in an One-Health approach, including (but not limited to) ecological surveillance of vegetation responses to climate change, land-use transformation and introduction of invasive species, and biodiversity shifts.

CRediT authorship contribution statement

Maria P. Plaza: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Jose Oteros:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Vivien Leier-Wirtz:** Writing – review & editing, Investigation, Data curation. **Athanasios Charalampopoulos:** Writing – review & editing, Investigation, Data curation. **Carmen Galán:** Writing – review & editing, Investigation, Conceptualization. **Caroline Holzmann:** Writing – review & editing, Investigation, Funding acquisition. **Franziska Kolek:** Writing – review & editing, Investigation, Data curation. **Despoina Vokou:** Writing – review & editing, Investigation, Data curation. **Claudia Traidl-Hoffmann:** Writing – review & editing, Investigation, Conceptualization. **Stefanie Gilles:** Writing – review & editing, Methodology, Investigation, Conceptualization. **Athanasios Damialis:** Writing – review & editing, Methodology, Investigation, Data curation, Conceptualization.

Funding sources

The study was funded by the German Federal Ministry of Research, Technology and Space (BMFTR) under the initiative *Junior Research Groups Climate, Environment and Health* (IMPACCT project – under funding code 01LN2202A). SG and CH was funded by the program KuHeMo (K1-2499-KuHeMo-21-V4) of the Bayerisches Landesamt für Gesunheit und Lebensmittelsicherheit (LGL).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envres.2026.124273.

Data availability

The authors do not have permission to share data.

References

Adamov, S., Pauling, A., 2023. A real-time calibration method for the numerical pollen forecast model COSMO-ART. Aerobiologia 39 (3), 327–344. https://doi.org/10.1007/s10453-023-09796-5.

Aguilera, F., Valenzuela, L.R., 2012. Microclimatic-induced fluctuations in the flower and pollen production rate of olive trees (Olea europaea L.). Grana 51 (3), 228–239. https://doi.org/10.1080/00173134.2012.659203.

Andersen, T.B., 1991. A model to predict the beginning of the pollen season. Grana 30 (1), 269–275. https://doi.org/10.1080/00173139109427810.

Annesi-Maesano, I., Cecchi, L., Biagioni, B., Chung, K.F., Clot, B., Collaud Coen, M., D'Amato, G., Damialis, A., Dominguez-Ortega, J., Galàn, C., Gilles, S., Holgate, S., Jeebhay, M., Kazadzis, S., Papadopoulos, N.G., Quirce, S., Sastre, J., Tummon, F., Traidl-Hoffmann, C., et al., 2023. Is exposure to pollen a risk factor for moderate and severe asthma exacerbations? Allergy. https://doi.org/10.1111/all.15724.

Astray, G., Amigo Fernández, R., Fernández-González, M., Dias-Lorenzo, D.A., Guada, G., Rodríguez-Rajo, F.J., 2025. Machine learning to forecast airborne parietaria pollen in the North-West of the iberian peninsula. Sustainability 17 (4), 1528. https://doi.org/10.3390/su17041528.

Beck, H.E., McVicar, T.R., Vergopolan, N., Berg, A., Lutsko, N.J., Dufour, A., Zeng, Z., Jiang, X., van Dijk, Miralles, D.G., 2023. High-resolution (1 km) Köppen-Geiger maps for 1901–2099 based on constrained CMIP6 projections. Scientific Data 10 (1), 1–16. https://doi.org/10.1038/s41597-023-02549-6.

Buters, J., Clot, B., Galán, C., Gehrig, R., Gilge, S., Hentges, F., O'Connor, D., Sikoparija, B., Skjoth, C., Tummon, F., Adams-Groom, B., Antunes, C.M., Bruffaerts, N., Çelenk, S., Crouzy, B., Guillaud, G., Hajkova, L., Seliger, A.K., Oliver, G., et al., 2024. Automatic detection of airborne pollen: an overview. Aerobiologia 40 (1), 13–37. https://doi.org/10.1007/s10453-022-09750-x.

Buters, J., Prank, M., Sofiev, M., Pusch, G., Albertini, R., Annesi-Maesano, I., Antunes, C., Behrendt, H., Berger, U., Brandao, R., others, 2015. Variation of the group 5 grass pollen allergen content of airborne pollen in relation to geographic location and time in season. J. Allergy Clin. Immunol. 136 (1), 87–95.

Cassagne, E., Caillaud, P.D., Besancenot, J.-P., Thibaudon, M., 2007. Forecasting the onset of an allergic risk to poaceae in Nancy and Strasbourg (France) with different methods. European Annals of Allergy and Clinical Immunology 39 (39), 262–268.

Cebrino, J., Galán, C., Domínguez-Vilches, E., 2016. Aerobiological and phenological study of the main Poaceae species in Córdoba City (Spain) and the surrounding hills. Aerobiologia 32 (4), 595–606. https://doi.org/10.1007/s10453-016-9434-6.

Cleveland, R., Cleveland, W., McRae, J., Terpenning, I., 1990. STL: a seasonal-trend decomposition procedure based on loess. J. Off. Stat. 6, 3–73. https://www.math.unm.edu/~lil/Stat581/STL.pdf.

Cordero, J.M., Rojo, J., Gutiérrez-Bustillo, A.M., Narros, A., Borge, R., 2021. Predicting the Olea pollen concentration with a machine learning algorithm ensemble. Int. J. Biometeorol. 65 (4), 541–554. https://doi.org/10.1007/s00484-020-02047-z.

D'Amato, G., Cecchi, L., Bonini, S., Nunes, C., Annesi-Maesano, I., Behrendt, H., Liccardi, G., Popov, T., van Cauwenberge, P., 2007. Allergenic pollen and pollen allergy in Europe. Allergy 62 (9), 976–990. https://doi.org/10.1111/j.1398-9995.2007.01393.x.

Damialis, A., Gioulekas, D., Lazopoulou, C., Balafoutis, C., Vokou, D., 2005. Transport of airborne pollen into the city of Thessaloniki: the effects of wind direction, speed and persistence. Int. J. Biometeorol. 49 (3), 139–145. https://doi.org/10.1007/s00484-004-0229-z.

Damialis, A., Traidl-Hoffmann, C., Treudler, R., 2019. Climate change and pollen allergies. In: Marselle, M.R., Stadler, J., Korn, H., Irvine, K.N., Bonn, A. (Eds.), Biodiversity and Health in the Face of Climate Change. Springer International Publishing, pp. 47–66. https://doi.org/10.1007/978-3-030-02318-8_3.

de Weger, L.A., Bergmann, K.C., Rantio-Lehtimäki, A., Dahl, Å., Buters, J., Déchamp, C., Belmonte, J., Thibaudon, M., Cecchi, L., Besancenot, J.-P., Galán, C., Waisel, Y., 2013. Impact of pollen. In: Sofiev, M., Bergmann, K.-C. (Eds.), Allergenic Pollen: a Review of the Production, Release, Distribution and Health Impacts. Springer, Netherlands, pp. 161–215. https://doi.org/10.1007/978-94-007-4881-1_6.

Devadas, R., Huete, A.R., Vicendese, D., Erbas, B., Beggs, P.J., Medek, D., Haberle, S.G., Newnham, R.M., Johnston, F.H., Jaggard, A.K., Campbell, B., Burton, P.K., Katelaris, C.H., Newbigin, E., Thibaudon, M., Davies, J.M., 2018. Dynamic ecological observations from satellites inform aerobiology of allergenic grass pollen. Sci. Total Environ. 633, 441–451. https://doi.org/10.1016/j.scitotenv.2018.03.191.

Emmerson, K.M., Silver, J.D., Thatcher, M., Wain, A., Jones, P.J., Dowdy, A., Newbigin, E.J., Picking, B.W., Choi, J., Ebert, E., Bannister, T., 2021. Atmospheric modelling of grass pollen rupturing mechanisms for thunderstorm asthma prediction. PLoS One 16 (4), e0249488. https://doi.org/10.1371/journal.pone.0249488.

Erbas, B., Chang, J.-H., Dharmage, S., Ong, E.K., Hyndman, R., Newbigin, E., Abramson, M., 2007. Do levels of airborne grass pollen influence asthma hospital admissions? Clin. Exp. Allergy: Journal of the British Society for Allergy and Clinical Immunology 37 (11), 1641–1647. https://doi.org/10.1111/j.1365-2222.2007.02818.x.

Fernández-Rodríguez, S., Durán-Barroso, P., Silva-Palacios, I., Tormo-Molina, R., Maya-Manzano, J.M., Gonzalo-Garijo, Á., 2016. Forecast model of allergenic hazard using trends of Poaceae airborne pollen over an urban area in SW Iberian Peninsula (Europe). Nat. Hazards 84 (1), 121–137. https://doi.org/10.1007/s11069-016-2411-0.

Galán, C., González, P.C., Teno, P.A., Vilches, E.D., 2007. Manual de calidad y gestión de la Red Española de Aerobiología. Universidad de Córdoba Córdoba. https://www.aerobiologia.com/app/download/5783625283/manual_cast.pdf.

Galán Díaz, J., Romero-Morte, J., Cascón, Á., Gutiérrez-Bustillo, A.M., Cervigón, P., Rojo, J., 2024. Different phenological behaviour of native and exotic grasses extends the period of pollen exposure with clinical implications in the Madrid Region, Spain. Biol. Invasions 26 (7), 2171–2182. https://doi.org/10.1007/s10530-024-03303-8.

García-Mozo, H., 2017. Poaceae pollen as the leading aeroallergen worldwide: a review. Allergy 72 (12), 1849–1858. https://doi.org/10.1111/all.13210.

García-Mozo, H., Yaezel, L., Oteros, J., Galán, C., 2014. Statistical approach to the analysis of olive long-term pollen season trends in southern Spain. Sci. Total Environ. 473–474, 103–109. https://doi.org/10.1016/j.scitotenv.2013.11.142.

Gehrig, R., Clot, B., 2021. 50 years of pollen monitoring in basel (Switzerland) demonstrate the influence of climate change on airborne pollen. Front. Allergy 2. https://www.frontiersin.org/articles/10.3389/falgy.2021.677159.

Goudarzi, G., Birgani, Y.T., Assarehzadegan, M.-A., Neisi, A., Dastoorpoor, M., Sorooshian, A., Yazdani, M., 2022. Prediction of airborne pollen concentrations by artificial neural network and their relationship with meteorological parameters and air pollutants. J. Environ. Health Sci. Eng. 20 (1), 251–264. https://doi.org/10.1007/s40201-021-00773-z.

Grewling, Ł., Jackowiak, B., Nowak, M., Uruska, A., Smith, M., 2012. Variations and trends of birch pollen seasons during 15 years (1996–2010) in relation to weather conditions in Poznań (western Poland). Grana 51 (4), 280–292. https://doi.org/10.1080/00173134.2012.700727.

Hernández-Ceballos, M.A., López-Orozco, R., Tenor-Ortiz, M.J., Galán, C., García-Mozo, H., 2026. Wind dynamics drives the changes of the 2001–2023 grass pollen seasons in Córdoba (southern Spain). Agric. For. Meteorol. 378, 110955. https://doi.org/10.1016/j.agrformet.2025.110955.

Hernandez-Ceballos, M.A., Soares, J., García-Mozo, H., Sofiev, M., Bolivar, J.P., Galán, C., 2014. Analysis of atmospheric dispersion of olive pollen in southern Spain using SILAM and HYSPLIT models. Aerobiologia 30 (3), 239–255. https://doi.org/10.1007/s10453-013-9324-0.

Hirst, J.M., 1952. An automatic volumetric spore trap. Ann. Appl. Biol. 39 (2), 257–265. https://doi.org/10.1111/j.1744-7348.1952.tb00904.x.

Holzmann, C., Karg, J., Reiger, M., Kharbal, R., Romano, P., Scheiwein, S., Khalfi, C., Muzalyova, A., Brunner, J.O., Hammel, G., Damialis, A., Traidl-Hoffmann, C., Plaza, M.P., Gilles, S., 2025. Clinical benefits of a randomized allergy app intervention in grass pollen sufferers: a controlled trial. Allergy (n/a). https://doi.org/10.1111/all.16558.

Hughes, D.D., Mampage, C.B.A., Jones, L.M., Liu, Z., Stone, E.A., 2020. Characterization of atmospheric pollen fragments during springtime thunderstorms. Environ. Sci. Technol. Lett. 7 (6), 409–414. https://doi.org/10.1021/acs.estlett.0c00213.

Hurtado, S., Antequera-Gómez, M.L., Barba-González, C., Picornell, A., Navas-Delgado, I., 2024. e-Science workflow: a semantic approach for airborne pollen prediction. Knowl. Base Syst. 284, 111230. https://doi.org/10.1016/j.knosys.2023.111230.

IPCC, 2022. Sixth assessment report of the intergovernmental panel on climate change. Working Group II: Impacts, Adaptation and Vulnerability. https://report.ipcc.ch/ar6wg2/pdf/IPCC_AR6_WGII_FinalDraft_FullReport.pdf.

Kuhn, M., 2008. Building predictive models in *R* using the **caret** package. J. Stat. Software 28 (5). https://doi.org/10.18637/jss.v028.i05.

Kurganskiy, A., Creer, S., de Vere, N., Griffith, G.W., Osborne, N.J., Wheeler, B.W., McInnes, R.N., Clewlow, Y., Barber, A., Brennan, G.L., Hanlon, H.M., Hegarty, M., Potter, C., Rowney, F., Adams-Groom, B., Petch, G.M., Pashley, C.H., Satchwell, J., de Weger, L.A., et al., 2021. Predicting the severity of the grass pollen season and the effect of climate change in Northwest Europe. Sci. Adv. 7 (13). https://doi.org/10.1126/sciadv.abd7658.

Liang, Z., Wei, X., Sun, X., 2022. An investigation into the variant-weight multi-model ensemble forecasting technique based on the analyses of model systematic errors calculation and elimination. Earth Space Sci. 9 (10). https://doi.org/10.1029/2022EA002547.

Liu, Shao, W., Lei, X., Shao, W., Gao, Z., Sun, J., Yang, S., Cai, Y., Ding, Z., Sun, N., Gu, S., Peng, L., Zhao, Z., 2025. Validation of the automatic real-time monitoring of airborne pollens in China against the reference hirst-type trap method. Atmosphere 16 (5), 531. https://doi.org/10.3390/atmos16050531.

Liu, X., Wu, D., Zewdie, G.K., Wijerante, L., Timms, C.I., Riley, A., Levetin, E., Lary, D.J., 2017. Using machine learning to estimate atmospheric Ambrosia pollen concentrations in Tulsa, OK. Environ. Health Insights 11, 1178630217699399. https://doi.org/10.1177/1178630217699399.

Lo, F., Bitz, C.M., Hess, J.J., 2021. Development of a Random Forest model for forecasting allergenic pollen in North America. Sci. Total Environ. 773, 145590. https://doi.org/10.1016/j.scitotenv.2021.145590.

López-Orozco, R., García-Mozo, H., Oteros, J., Galán, C., 2023. Long-term trends and influence of climate and land-use changes on pollen profiles of a Mediterranean oak forest. Sci. Total Environ. 897, 165400. https://doi.org/10.1016/j.scitotenv.2023.165400.

Makra, L., Coviello, L., Gobbi, A., Jurman, G., Furlanello, C., Brunato, M., Ziska, L.H., Hess, J.J., Damialis, A., Garcia, M.P.P., Tusnády, G., Czibolya, L., Ihász, I., Deák, Á. J., Mikó, E., Dorner, Z., Harry, S.K., Bruffaerts, N., Packeu, A., et al., 2024. Forecasting daily total pollen concentrations on a global scale. Allergy 79 (8), 2173–2185. https://doi.org/10.1111/all.16227.

Martínez-Bracero, M., García-Llamas, C., López-Orozco, R., Oteros, J., Alcázar, P., Galán, C., 2025. Environmental changes on grass flowering phenological trends (2000–2021). Plants, People, Planet 7 (6), 1824–1833. https://doi.org/10.1002/ppp3.70060.

Maya-Manzano, J.M., Tummon, F., Abt, R., Allan, N., Bunderson, L., Clot, B., Crouzy, B., Daunys, G., Erb, S., Gonzalez-Alonso, M., Graf, E., Grewling, Ł., Haus, J., Kadantsev, E., Kawashima, S., Martinez-Bracero, M., Matavulj, P., Mills, S., Niederberger, E., et al., 2023. Towards European automatic bioaerosol monitoring: Comparison of 9 automatic pollen observational instruments with classic Hirst-type traps. Sci. Total Environ. 866, 161220. https://doi.org/10.1016/j.scitotenv.2022.161220.

Milic, A., Addison-Smith, B., Jones, P.J., Beggs, P.J., Erbas, B., Davies, J.M., 2020. Quality control of pollen identification and quantification exercise for the AusPollen

Aerobiology Collaboration Network: a pilot study. Aerobiologia 36 (1), 83–87. https://doi.org/10.1007/s10453-019-09580-4.
Muzalyova, A., Brunner, J.O., Traidl-Hoffmann, C., Damialis, A., 2021. Forecasting Betula and Poaceae airborne pollen concentrations on a 3-hourly resolution in Augsburg, Germany: toward automatically generated, real-time predictions. Aerobiologia 37 (3), 425–446. https://doi.org/10.1007/s10453-021-09699-3.
Navares, R., Aznarte, J.L., 2020. Predicting air quality with deep learning LSTM: towards comprehensive models. Ecol. Inform. 55, 101019. https://doi.org/10.1016/j.ecoinf.2019.101019.
Nickovic, S., Petković, S., Ilić, L., Pejanović, G., Mijić, Z., Huete, A., Marks, G., 2023. Prediction of airborne pollen and sub-pollen particles for thunderstorm asthma outbreaks assessment. Sci. Total Environ. 864, 160879. https://doi.org/10.1016/j.scitotenv.2022.160879.
Oh, J.-W., 2022. Pollen allergy in a changing planetary environment. Allergy Asthma Immunol. Res. 14 (2), 168–181. https://doi.org/10.4168/aair.2022.14.2.168.
Oteros, J., García-Mozo, H., Hervás, C., Galán, C., 2013. Biometeorological and autoregressive indices for predicting olive pollen intensity. Int. J. Biometeorol. 57 (2), 307–316. https://doi.org/10.1007/s00484-012-0555-5.
Oteros, J., Orlandi, F., García-Mozo, H., Aguilera, F., Dhiab, A.B., Bonofiglio, T., Abichou, M., Ruiz-Valenzuela, L., del Trigo, M.M., Díaz de la Guardia, C., Domínguez-Vilches, E., Msallem, M., Fornaciari, M., Galán, C., 2014. Better prediction of Mediterranean olive production using pollen-based models. Agron. Sustain. Dev. 34 (3), 685–694. https://doi.org/10.1007/s13593-013-0198-x.
Pfaar, O., Karatzas, K., Bastl, K., Berger, U., Buters, J., Darsow, U., Demoly, P., Durham, S.R., Galán, C., Gehrig, R., Gerth van Wijk, R., Jacobsen, L., Katsifarakis, N., Klimek, L., Saarto, A., Sofiev, M., Thibaudon, M., Werchan, B., Bergmann, K.-C., 2020. Pollen season is reflected on symptom load for grass and birch pollen-induced allergic rhinitis in different geographic areas—An EAACI Task Force Report. Allergy 75 (5), 1099–1106. https://doi.org/10.1111/all.14111.
Picornell, A., Hurtado, S., Antequera-Gómez, M.L., Barba-González, C., Ruiz-Mata, R., de Gálvez-Montañez, E., Recio, M., Trigo, M. del M., Aldana-Montes, J.F., Navas-Delgado, I., 2024. A deep learning LSTM-based approach for forecasting annual pollen curves: *olea* and Urticaceae pollen types as a case study. Comput. Biol. Med. 168, 107706. https://doi.org/10.1016/j.compbiomed.2023.107706.
Picornell, A., Ruiz-Mata, R., Rojo, J., Oteros, J., Recio, M., de Gálvez-Montañez, E., Trigo, M.M., 2023. Applying wind patterns and land use to estimate the concentrations of airborne pollen of herbaceous taxa in a statistical framework. Urban Clim. 49, 101496. https://doi.org/10.1016/j.uclim.2023.101496.
Plaza, M.P., Kolek, F., Leier-Wirtz, V., Brunner, J.O., Traidl-Hoffmann, C., Damialis, A., 2022. Detecting airborne pollen using an automatic, real-time monitoring system: evidence from two sites. Int. J. Environ. Res. Publ. Health 19 (4). https://doi.org/10.3390/ijerph19042471. Article 4.
Posit team, 2023. Rstudio: Integrated Development Environment for R. PBC, Posit. http://www.posit.co.
R Core Team, 2022. R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing. https://www.R-project.org.
Rodriguez-Rajo, F.J., Valencia-Barrera, R.M., Vega-Maray, A.M., Suarez, F.J., Fernandez-Gonzales, D., Jato, V., 2006. Prediction of airborne Alnus pollen concentration by using ARIMA models. Ann. Agric. Environ. Med. 13 (1). http://agro.icm.edu.pl/agro/element/bwmeta1.element.agro-article-6dbb9034-d274-4b68-a054-e0564446fcf7.
Rojo, J., Oteros, J., Picornell, A., Maya-Manzano, J.M., Damialis, A., Zink, K., Werchan, M., Werchan, B., Smith, M., Menzel, A., Timpf, S., Traidl-Hoffmann, C., Bergmann, K.-C., Schmidt-Weber, C.B., Buters, J., 2021. Effects of future climate change on birch abundance and their pollen load. Glob. Change Biol. 27 (22), 5934–5949. https://doi.org/10.1111/gcb.15824.
Rojo, J., Pérez-Badia, R., 2015. Models for forecasting the flowering of Cornicabra olive groves. Int. J. Biometeorol. 59 (11), 1547–1556.
Rojo, J., Picornell, A., Oteros, J., 2019. AeRobiology: the computational tool for biological data in the air. Methods Ecol. Evol. 10 (8), 1371–1376. https://doi.org/10.1111/2041-210X.13203.
Rojo, J., Romero-Morte, J., Lara, B., Quirós, E., Richardson, A.D., Pérez-Badia, R., 2022. Biological-based and remote sensing techniques to link vegetative and reproductive development and assess pollen emission in Mediterranean grasses. Ecol. Inform. 72, 101898. https://doi.org/10.1016/j.ecoinf.2022.101898.
Rojo, J., Salido, P., Pérez-Badia, R., 2015. Flower and pollen production in the 'cornicabra'olive (Olea europaea L.) cultivar and the influence of environmental factors. Trees (Berl.) 29 (4), 1235–1245.
Ruan, W., Li, Z., Sun, Z., An, X., Zhao, Y., Zhang, S., Liang, Y., Bu, Y., Xin, J., Hang, X., 2024. Enhancing pollen prediction in Beijing, a Chinese megacity: leveraging ensemble learning models for greater accuracy. Aerosol Air Qual. Res. 24 (11), 240123. https://doi.org/10.4209/aaqr.240123.
Siljamo, P., Sofiev, M., Filatova, E., Grewling, Ł., Jäger, S., Khoreva, E., Linkosalo, T., Ortega Jimenez, S., Ranta, H., Rantio-Lehtimäki, A., Svetlov, A., Veriankaite, L., Yakovleva, E., Kukkonen, J., 2013. A numerical model of birch pollen emission and dispersion in the atmosphere. Model evaluation and sensitivity analysis. Int. J. Biometeorol. 57 (1), 125–136. https://doi.org/10.1007/s00484-012-0539-5.
Silver, J.D., Spriggs, K., Haberle, S.G., Katelaris, C.H., Newbigin, E.J., Lampugnani, E.R., 2020. Using crowd-sourced allergic rhinitis symptom data to improve grass pollen forecasts and predict individual symptoms. Sci. Total Environ. 720, 137351. https://doi.org/10.1016/j.scitotenv.2020.137351.
Smith, M., Oteros, J., Schmidt-Weber, C., Buters, J.T.M., 2019. An abbreviated method for the quality control of pollen counters. Grana 58 (3), 185–190. https://doi.org/10.1080/00173134.2019.1570327.
Sofiev, M., Berger, U., Prank, M., Vira, J., Arteta, J., Belmonte, J., Bergmann, K.-C., Chéroux, F., Elbern, H., Friese, E., 2015. MACC regional multi-model ensemble simulations of birch pollen dispersion in Europe. Atmos. Chem. Phys. 15 (14), 8115–8130.
Sofiev, M., Ritenberga, O., Albertini, R., Arteta, J., Belmonte, J., Bernstein, C.G., Bonini, M., Celenk, S., Damialis, A., Douros, J., Elbern, H., Friese, E., Galan, C., Oliver, G., Hrga, I., Kouznetsov, R., Krajsek, K., Magyar, D., Parmentier, J., et al., 2017. Multi-model ensemble simulations of olive pollen distribution in Europe in 2014: current status and outlook. Atmos. Chem. Phys. 17 (20), 12341–12360. https://doi.org/10.5194/acp-17-12341-2017.
Stevanovic, K., Sinkkonen, A., Pawankar, R., Zuberbier, T., 2025. Urban greening and pollen allergy: balancing health and environmental sustainability. J. Allergy Clin. Immunol. Pract. 13 (2), 275–279. https://doi.org/10.1016/j.jaip.2024.12.017.
Suanno, C., Aloisi, I., Fernández-González, D., Del Duca, S., 2021. Pollen forecasting and its relevance in pollen allergen avoidance. Environ. Res. 200, 111150. https://doi.org/10.1016/j.envres.2021.111150.
Valipour Shokouhi, B., de Hoogh, K., Gehrig, R., Eeftens, M., 2024a. Estimation of historical daily airborne pollen concentrations across Switzerland using a spatio temporal random forest model. Sci. Total Environ. 906, 167286. https://doi.org/10.1016/j.scitotenv.2023.167286.
Valipour Shokouhi, B., de Hoogh, K., Gehrig, R., Eeftens, M., 2024b. Spatiotemporal modelling of airborne birch and grass pollen concentration across Switzerland: a comparison of statistical, machine learning and ensemble methods. Environ. Res. 263, 119999. https://doi.org/10.1016/j.envres.2024.119999.
Voukantsis, D., Niska, H., Karatzas, K., Riga, M., Damialis, A., Vokou, D., 2010. Forecasting daily pollen concentrations using data-driven modeling methods in Thessaloniki, Greece. Atmos. Environ. 44 (39), 5101–5111. https://doi.org/10.1016/j.atmosenv.2010.09.006.
Wang, J., Yang, Y., Gong, X., 2024. Interpretable machine learning for allergic rhinitis prediction among preschool children in Urumqi, China. Sci. Rep. 14, 22281. https://doi.org/10.1038/s41598-024-73733-w.
World Meteorological Organization, 2017. WMO Guidelines on the Calculation of Climate Normals. Geneva, Switzerland. ISBN 978-92-63-11203-3. https://www.agroorbi.pt/livroagrometeorologia/DocsProg/Temas&Exerc%C3%ADciosExtraPorCap%C3%ADtulo/Cap1_Introdu%C3%A7%C3%A3o/Docs/WMO%20Guidelines%20on%20the%20Calculation%20of%20Climate%20Normals_en.pdf.
Yin, J., Zhang, Y., Du, Y., Ouyang, Y., Wang, C., Ma, Z., Wang, H., Sun, S., Zhang, L., Chen, R., 2026. High-resolution mapping of allergenic pollen risk across China using ensemble machine learning. Ecotoxicol. Environ. Saf. 309, 119659. https://doi.org/10.1016/j.ecoenv.2025.119659.
Zewdie, G.K., Lary, D.J., Levetin, E., Garuma, G.F., 2019. Applying deep neural networks and ensemble machine learning methods to forecast airborne Ambrosia pollen. Int. J. Environ. Res. Publ. Health 16 (11), 1992. https://doi.org/10.3390/ijerph16111992.
Zhu, X., Ma, X., Zhang, Z., Liu, Y., Luo, Y., Yan, K., Pei, T., Huete, A., 2024. Floating in the air: forecasting allergenic pollen concentration for managing urban public health. Int. J. Digit. Earth 17 (1), 2306894. https://doi.org/10.1080/17538947.2024.2306894.