Time Series – Heat-Pollution Index for Pollen Allergy Risk

Subject: Heat-Pollution Index (HPI): does a heat × pollution composite improve pollen allergy risk modelling? A multi-city study across Paris, Marseille, Strasbourg and Bordeaux (2022–2024).
Authors: Jonathan Bouniol, Anna Spira, Sacha Nardoux, Enzo Natali, Keira Chang.
Date: 9 May 2026.
Instructor: Firas Bilbeche.
Code repository: github.com/jbouniol/time-series-pollen.

Table of Content

Table of Content
1. Introduction
2. Hypotheses
3. Data
    3.1 Data sources
    3.2 Data preparation
    3.3 Construction of the Heat-Pollution Index (HPI)
4. Exploratory Data Analysis
    4.1 Pollen: distributions and seasonality
    4.2 HPI: temporal behaviour and inter-city contrasts
    4.3 Historical evolution: why the HPI will worsen
    4.4 Ameli: a semi-annual health signal
5. Time series methodology
    5.1 Justification of the model progression
    5.2 Stationarity and transformation
    5.3 Architecture of the three model families
    5.4 Evaluation protocol
6. Results
    6.1 Establishing the baseline: pollen-only model performance
    6.2 H1: residuals reveal what the pollen-only model misses
    6.3 H2 and H3: the HPI improves prediction, differently across cities
    6.4 H4: consistency with Ameli health data
7. Discussion and limitations
    7.1 What the results allow us to claim
    7.2 Limitations
    7.3 Avenues for extension
8. Conclusion
9. References
    Scientific articles
    Reference event
    Institutional sources
    Data sources
    Software tools and libraries

1. Introduction

Pollen allergies are one of the major environmental pathologies of industrialised countries, and their prevalence keeps rising. In France, the Ministry of Health estimates that approximately **30% of adults** and 20% of children over nine suffer from them, against a much smaller share thirty years ago. At the European scale, the European Environment Agency reports that **40% of the population** is affected by allergic rhinitis, allergic asthma or rhinoconjunctivitis. Beyond clinical symptoms, these allergies are costly: degraded sleep, reduced productivity, lower academic performance, increased healthcare use.

Against this backdrop, current forecasting systems (RNSA bulletins in France, European monitoring networks) rely on **a single signal**: airborne pollen grain concentration, expressed in grains per cubic metre. The metric is convenient and historically validated, but the continued growth in allergic prevalence raises a simple question: if pollen exposure alone

explained the risk, why would the number of affected people keep increasing at this pace? Something in our environment, or in our way of life, seems to amplify pollen's effect on the population.

The 2016 Melbourne thunderstorm asthma episode (more than 3 300 ER visits and ten deaths in thirty hours, for grass pollen concentrations comparable to other days without incident) concretely illustrates this idea: **at equal pollen load, the health risk can vary substantially with environmental context**. Recent literature points to two main suspects. Tropospheric ozone (O_3), studied notably by D'Amato. (2020), would chemically modify pollen grains and increase their biological aggressiveness. Nitrogen dioxide (NO_2), tracked by Sousa-Silva et al. (2021), would facilitate allergen penetration into the airways. Crucially, both pollutants are tightly coupled to **heat**: O_3 forms through photochemical reactions that intensify with temperature, and NO_2 accumulates during hot, stagnant episodes. Heat and pollution therefore likely do not act independently, they may reinforce one another.

This is the idea we put to the test in this work. The central research question is:

> **Does pollen concentration alone suffice to characterise allergic risk, or does a composite index combining heat and pollution carry additional, measurable information consistent with observed health data?**

To answer it, we build a **Heat-Pollution Index (HPI)**, a composite combining temperature and atmospheric pollution, and evaluate it on four French cities with contrasted climatic profiles: **Paris** (urban semi-oceanic), **Marseille** (Mediterranean), **Strasbourg** (continental) and **Bordeaux** (oceanic). The study pursues four interconnected objectives:

- **Methodological**: diagnose whether pollen forecasting models miss an environmental signal.
- **Temporal**: model that signal at daily resolution over 2022–2024 via time series.
- **Spatial**: leverage the contrast between the four cities to test whether the effect depends on the urban profile.
- **Sanitary**: confront the resulting signal with an external dataset, antihistamine reimbursements published by the French health insurance system (Assurance Maladie).

The remainder of the report is organised as follows. Section 2 breaks the question into four testable hypotheses. Section 3 describes the data sources and HPI construction. Section 4 presents the exploratory analysis. Section 5 justifies the time series methodology. Section 6 lays out the results in four acts. Sections 7 and 8 discuss limitations and conclude.

2. Hypotheses

The research question raises four distinct empirical questions, which we formulate as four testable hypotheses. Each maps to one act of the analysis presented in Section 6.

H1: The pollen-only model is incomplete.

> *The days on which a model trained on pollen alone makes the largest errors are systematically days of high heat and high pollution.*

If forecasting errors concentrate on episodes of combined environmental stress, the model is ignoring a relevant explanatory variable, in other words, it is misspecified. This is the direct motivation for adding a heat × pollution piece of information.

H2: The HPI carries useful information.

> *Including the HPI as an exogenous regressor improves pollen forecasting compared to a model using only raw weather variables.*

Concretely, if the HPI captures something that temperature or pollution taken separately do not, then a model that includes it must predict better than one that does not. Three model families are compared: weather only (M1), HPI only (M2), weather + HPI (M3).

H3: The effect depends on the urban profile.

> *The gain from the HPI is more pronounced in cities with chronic pollution (Paris, Marseille) than in cities with lower background pollution (Bordeaux).*

If allergenicity amplification is genuinely driven by pollution, its intensity must follow chronic urban exposure. A uniform gain across the four cities would, conversely, suggest the HPI captures a generic seasonal effect rather than a specific mechanism.

H4: The signal shows up in health data.

> *Semesters with high HPI correspond to higher antihistamine reimbursements, and this relationship also holds in 2014.*

If the HPI reflects a real allergic risk, its variations should at least directionally show up in population-level consumption of ATC class R06A (systemic antihistamines). The year 2014, outside the main 2022–2024 window, serves as a temporal anchor verifying that the relationship is not a recent artefact. Ameli's semi-annual granularity calls for reading the result as **directional consistency** rather than strict causal demonstration.

These four hypotheses read in order: H1 shows the standard model is insufficient, H2 proposes and tests an alternative, H3 verifies it behaves consistently across cities, and H4 confronts the result against an independent health metric.

3. Data

3.1 Data sources

Four data sources are used in this study: pollen, weather, air quality, and drug reimbursements. They cover the same territory (the four cities studied) over periods adapted to the questions at hand.

Pollen and **air quality data** come from the Open-Meteo Air Quality API, backed by the Copernicus Atmosphere Monitoring Service (CAMS) ensemble model. Six pollen species were retained: birch (*Betula*), alder (*Alnus*), grasses (*Poaceae*), olive (*Olea*), mugwort (*Artemisia*) and ragweed (*Ambrosia*), along with three atmospheric pollutants: ozone (O_3), nitrogen dioxide (NO_2) and fine particles ($PM_{2.5}$). Data are available at hourly resolution and were aggregated into daily averages.

Weather data come from the Open-Meteo Historical Weather API, over 1980–2024, to feed the historical-trend analysis (Section 4.3). Retained variables are temperature (mean, max, min), cumulative precipitation, maximum wind speed, sunshine duration and relative humidity. This API leverages ERA5-Land reanalyses, providing homogeneous spatial coverage across the four cities, an advantage over ground-station networks, whose density varies sharply between cities.

Healthcare consumption data come from the Assurance Maladie open-data portal (Open Medic). Antihistamine reimbursements (ATC class R06A) are published at **monthly granularity** by region. Four target regions were retained, matching the cities studied: Île-de-France (Paris), PACA (Marseille), Grand Est (Strasbourg) and Nouvelle-Aquitaine (Bordeaux). The period covered is 2022–2024, supplemented by 2014 as a long-term temporal anchor.

Source	Variables	Granularity	Period
Open-Meteo Air Quality API	6 pollen species, O_3, NO_2, $PM_{2.5}$	Hourly → daily	2022–2024
Open-Meteo Historical Weather API	T° (mean/max/min), precip., wind, sunshine, humidity	Daily	1980–2024
Open Medic (Assurance Maladie)	Antihistamine reimbursements (R06A)	Monthly × region	2014, 2022–2025

Links to the APIs and the open-data portal are listed in the references.

3.2 Data preparation

Pollen, weather and air quality pipeline. The pollen and air quality series, available at hourly resolution, were resampled into daily averages. For pollen, missing values were imputed to zero, a biologically justified choice, since gaps systematically correspond to off-season periods where pollen absence is the most likely state. For weather and air quality, missing values were filled by linear interpolation within a maximum window of three consecutive days. The three streams were merged on the (date × city) key, building a **balanced panel of 4 cities × 1 096 daily observations** (1 January 2022 – 31 December 2024).

Ameli pipeline. Open Medic data are published at a monthly cadence by region and ATC3 class. The R06A class (systemic antihistamines) was extracted, and raw volumes were converted to **boxes per 1 000 inhabitants** using INSEE population estimates, neutralising demographic gaps between regions (without this normalisation, Île-de-France mechanically dominates every comparison). The final monthly panel covers **36 months over 2022–2024**, supplemented by **12 months of 2014** as a long-term temporal anchor: 192 points in total.

A caveat: consumption is observed at the regional level, not departmental, let alone urban. This granularity, broader than that of the other data layers, is inherent to the Open Medic publication and is an assumed limitation of the Ameli set-up (Section 7).

3.3 Construction of the Heat-Pollution Index (HPI)

The Heat-Pollution Index (HPI) is built as a multiplicative composite of thermal stress and atmospheric pollution. For city c on day t, it is defined as:

$$\mathrm{HPI}_{c,t} = T^{*}_{c,t} \times \frac{O^{*}_{3,c,t} + \mathrm{NO}^{*}_{2,c,t}}{2}$$

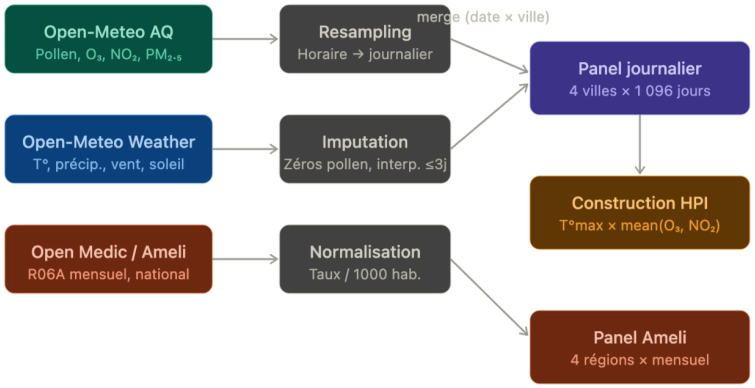
where T^*, O_3^* and NO_2^* denote the Min-Max normalised values, computed independently per city, of daily maximum temperature, ozone and nitrogen dioxide:

$$X_{c,t}^* = \frac{X_{c,t} - \min_t(X_c)}{\max_t(X_c) - \min_t(X_c)}$$

Normalisation is applied city by city in order to preserve inter-city contrasts while keeping the index bounded in [0,1]. A value close to 1 characterises a day of simultaneously high thermal stress **and** pollution, relative to the city's own history.

The choice of a **multiplicative** formulation is deliberate. An additive composite would assign a high score to a day of high pollution without heat, or of high heat without pollution. The product, by contrast, requires that **both conditions be met at the same time**. This is precisely the amplification effect we want to isolate: pollution and heat reinforcing one another on pollen allergenicity, rather than each acting on its own.

HPI variants were tested (additive formulation, addition of $PM_{2.5}$) and are presented in the appendix; the multiplicative O_3–NO_2 specification retained here is the one that best aligns with the modelling results in Section 6.



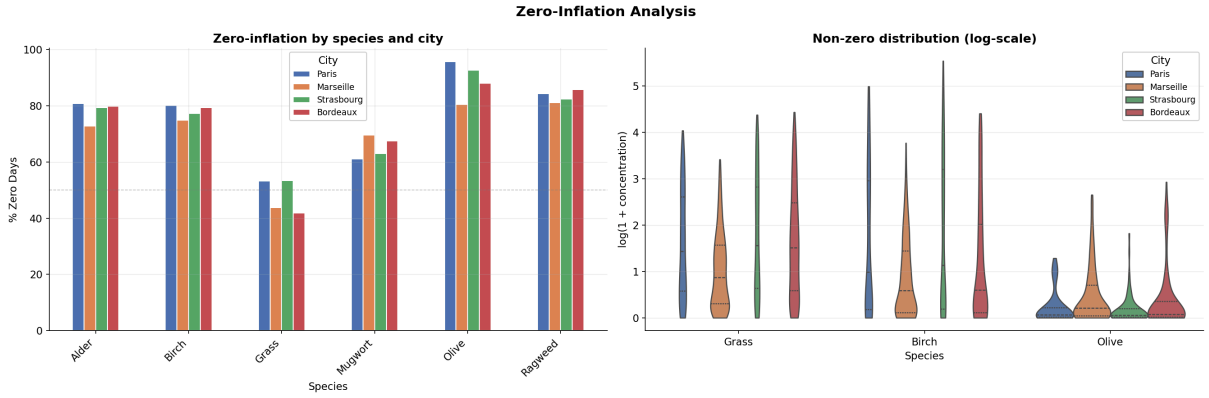
Data collection and preparation pipeline. The three source streams (Open-Meteo Air Quality, Open-Meteo Historical Weather, Open Medic/Ameli) are processed independently before being merged on the (date × city) key to form the daily panel (4 cities × 1,096 days) and the monthly Ameli panel. The HPI is built last from the normalised variables of the daily panel.

4. Exploratory Data Analysis

Before any modelling, we explore the four data layers that feed the analysis: pollen concentrations, the HPI composite, long-term climatic and air-quality trends, and the Ameli antihistamine signal. Each layer is examined for distribution, seasonality, and inter-city contrasts, the structural properties our hypotheses depend on.

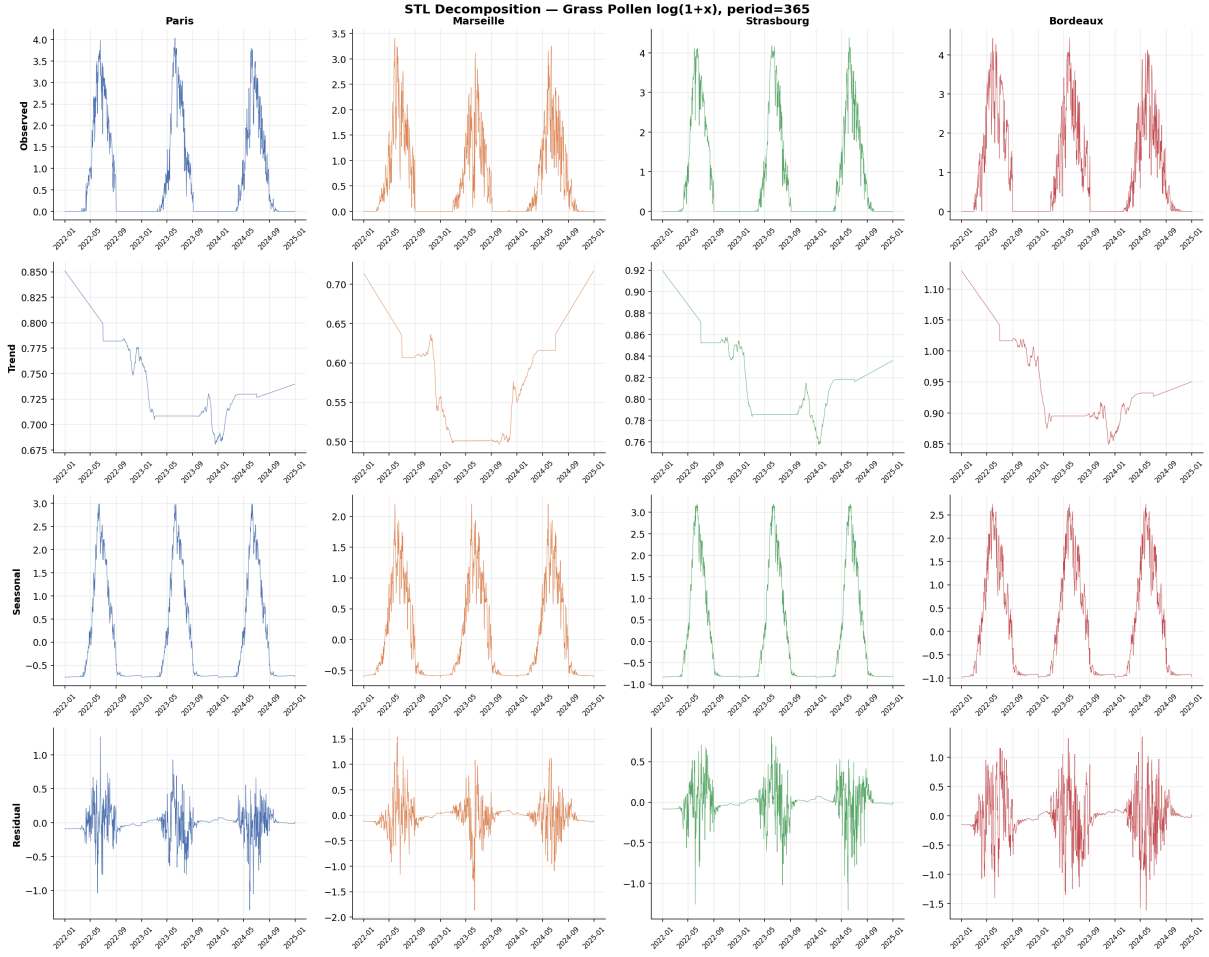
4.1 Pollen: distributions and seasonality

Pollen series are dominated by two structural features that shape every downstream modelling choice. First, **zero-inflation is severe**: between 42% (Marseille) and 96% (Strasbourg, ragweed) of daily values are zero, reflecting strict biological phenological windows. Second, the conditional distribution of non-zero days is **massively right-skewed** with skewness $\gamma_1 \in [2.9, 19.2]$, driven by short, intense release peaks. These properties violate the variance-stability assumption underlying ARIMA. We therefore apply the variance-stabilising transform $y_t = \log(1 + x_t)$ throughout the modelling pipeline, which compresses extreme peaks while preserving the zero structure.



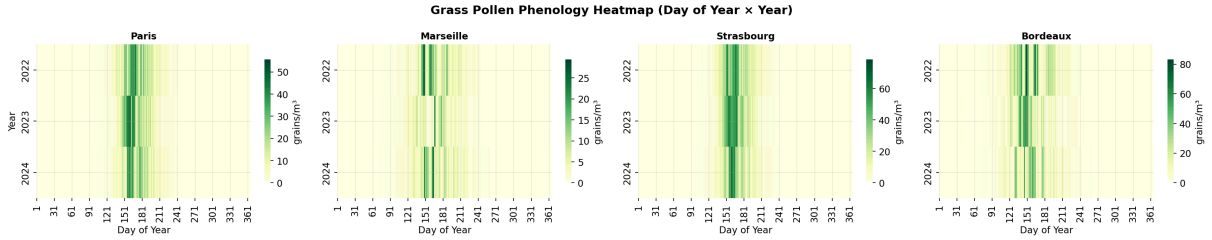
Share of zero days by city × species : *zero-inflation reaches 96% (Strasbourg, ragweed) and alone justifies the departure from the Gaussian ARIMA framework.*

STL decomposition expresses each series as $x_t = T_t + S_t + R_t$ (trend, seasonal, remainder) and confirms a dominant annual cycle for all six species. Phenological signatures are species-specific: alder and birch peak in late winter / early spring, grass dominates May–July, olive concentrates in Marseille (May–June), and ragweed produces a sharp late-summer pulse particularly pronounced in Strasbourg.



STL decomposition of grass pollen : *the annual seasonal component dominates, the residual stays centered but heteroscedastic around the peaks.*

A phenological heatmap (`02_eda.ipynb`) synthesises this calendar across all city × species combinations.



Average phenological calendar : *alder/birch in late winter, grass May–July, olive concentrated in Marseille, ragweed as a late pulse over Strasbourg.*

A **Seasonal Pollen Index** $\mathrm{SPIn}_{c,s,y} = \sum_{t \in \text{season}} x_{c,t}$ is computed by city c, species s, and year y to quantify inter-annual variability. A Kruskal-Wallis test confirms statistically significant inter-city differences for **5 of the 6 species** ($p < 0.01$), birch being the exception, consistent with its broader geographical homogeneity. This validates the four-city design as a genuinely contrasted natural experiment.

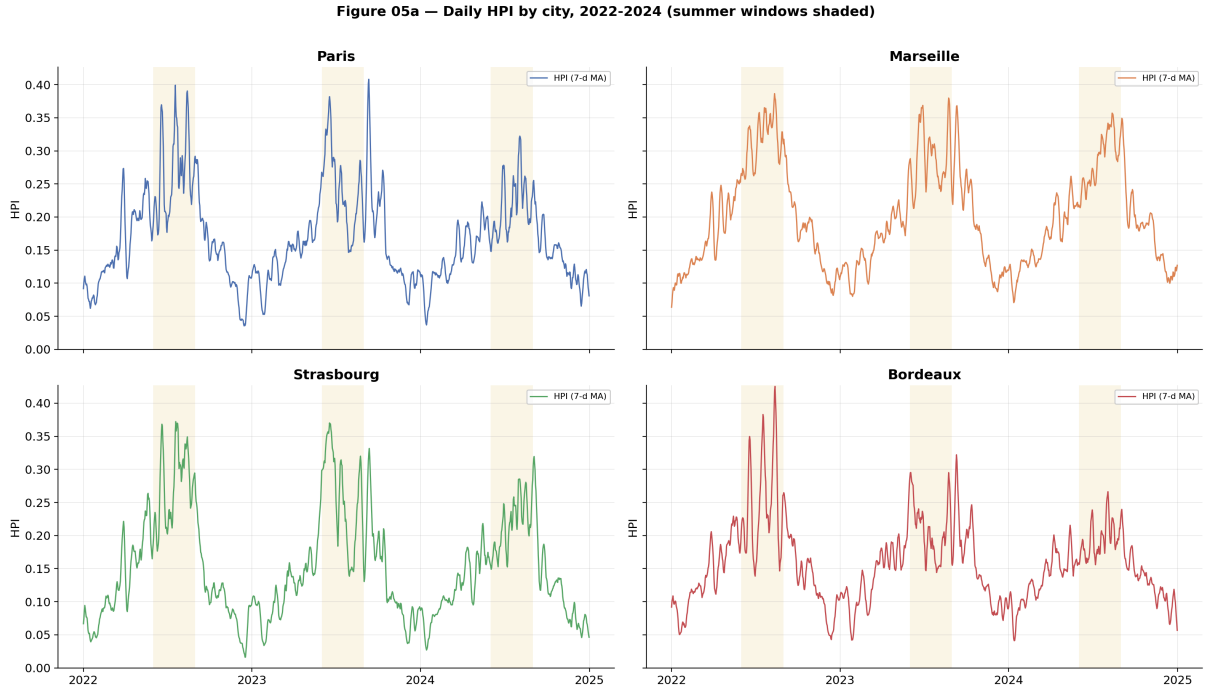
4.2 HPI: temporal behaviour and inter-city contrasts

The HPI is defined as:

$$\mathrm{HPI}_t = \tilde{T}_t \cdot \frac{\widetilde{O_{3t}} + \widetilde{\mathrm{NO}_{2t}}}{2}$$

where $\tilde{\cdot}$ denotes MinMax normalisation computed city-wise over the full 2022–2024 period. The index is bounded in $[0, 1]$, with high values indicating simultaneous thermal stress and atmospheric pollution.

The HPI time series (`05_hpi_analysis.ipynb`) exhibits a clean annual cycle peaking in **JJA** across all four cities, confirming that the heat × pollution interaction is structurally a summer phenomenon, the season when both ozone photochemical production and temperature are simultaneously elevated. Monthly profiles show June–August values 3–5× above winter baselines.

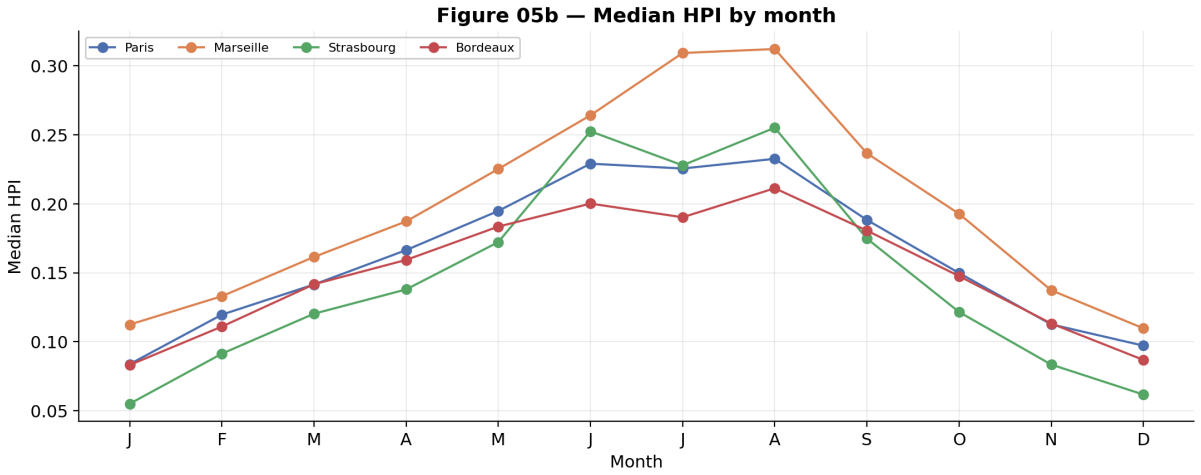


HPI series 2022–2024 by city : *clean annual cycle, summer peaks, summer 2022 visibly the most extreme across the four stations.*

Comparing the three summers, **2022 stands out** as the most extreme HPI year across all cities, consistent with the documented June–August 2022 European heatwave coinciding with elevated tropospheric ozone.

Crucially, the city ranking provides a **sanity check**: Paris > Strasbourg > Bordeaux ≈ Marseille in mean annual HPI. Paris's lead is driven by chronic NO_2 from urban traffic combining with summer heat. Importantly, city-wise normalisation does

not suppress this contrast: Paris's $\widetilde{O_3}$ and $\widetilde{NO_2}$ ceilings are higher in absolute terms, so its HPI median remains structurally above Bordeaux's.

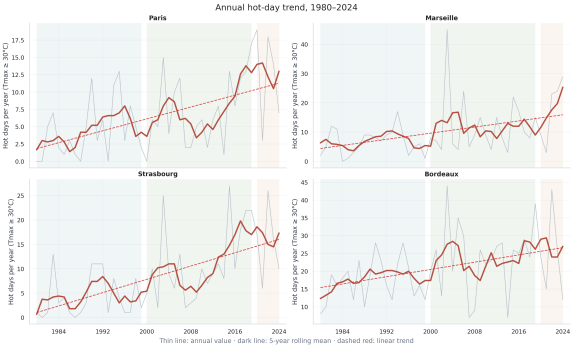


Average monthly profiles : *Paris stands out structurally in summer, chronic urban NO₂ amplifying the thermal component despite city-wise normalisation.*

4.3 Historical evolution: why the HPI will worsen

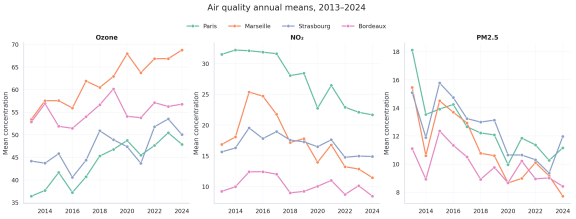
To argue that the HPI is not a 3-year curiosity but a **structurally rising risk**, `06_historical_evolution.ipynb` examines each component over its full available history.

Climate (1980–2024): Maximum temperatures have risen by $+1.4^\circ C$ to $+1.9^\circ C$ depending on the city; days above $30^\circ C$ have roughly doubled in Paris and Bordeaux and tripled in Marseille; frost days have collapsed across all cities. These trends are consistent with a warming rate of approximately $+0.3^\circ C$ per decade over the period.



Number of days > 30 °C per year : *doubling in Paris/Bordeaux, tripling in Marseille, the thermal substrate of the HPI thickens.*

Air quality (2013–2024): Ozone exhibits an upward trend in all four cities, driven by increased photochemical production on hotter days, a non-linear feedback, since ozone formation scales with both temperature and NO_x precursors. NO_2 shows a marginal decline in Paris and Bordeaux (vehicle-fleet transition) but remains structurally elevated.



Annual evolution of O₃ and NO₂ : *ozone rising everywhere (photochemical feedback), NO₂ slightly declining in Paris/Bordeaux but structurally high.*

The key insight is algebraic: if $\tilde{T}$ grows at rate α and $\frac{\widetilde{O_3}+\widetilde{NO_2}}{2}$ grows at rate β, the HPI grows approximately at rate $\alpha + \beta + \alpha\beta > \alpha + \beta$. **The interaction term grows super-additively** relative to its components. We do not need 30 years of pollen data to establish this: the two ingredients of the index are themselves trending upward, and their product amplifies the joint increase.

4.4 Ameli: a semi-annual health signal

Ameli reimbursements, normalised per 1 000 inhabitants to neutralise demographic bias, constitute a population-level proxy for the allergic risk actually experienced.

Two raw patterns emerge before any formal test. First, the spring months (April–June) systematically concentrate the consumption peaks across all four cities, consistent with the pollen calendar dominated by birch and grass. Second, the inter-city ranking of reimbursements broadly reflects that of the HPI, with Paris and Strasbourg leading per capita consumption.

A preliminary visualisation cross-references monthly reimbursements with the average monthly HPI, facetted by city. A positive directional association is visible even at comparable pollen levels: months with high HPI tend to correspond to higher reimbursements, independently of the raw pollen load. The formal regression is presented in section 6.3.

5. Time series methodology

5.1 Justification of the model progression

The modelling strategy follows a stepwise logic constrained by the empirical properties of the data: each model addresses a limit identified in the previous one, and each transition is justified either by an EDA result or a statistical test.

- **From the Seasonal Naïve to ARIMA.** The seasonal naïve model is the absolute baseline: it simply predicts the value observed on the same day a year earlier $\hat{x}_{t+h} = x_{t+h-365}$. While it exploits the annual periodicity of pollen series, it ignores any short-term temporal dependence. The Ljung-Box tests run in `02_eda.ipynb` reveal significant residual autocorrelation after the $log(1+x)$ transform, justifying the move to an ARIMA model capable of capturing this serial memory through its autoregressive and moving-average components.
- **From ARIMA to ARIMA+Fourier.** The natural extension of ARIMA for series with annual seasonality would be a $SARIMA(p,d,q)(P,D,Q)_{365}$. This approach is infeasible here: with only 730 training observations (2022–2023), a seasonal model with $s = 365$ would have to estimate several hundred seasonal parameters, far exceeding the dataset's capacity. The chosen solution is to model seasonality deterministically through Fourier regressors: K pairs of sines/cosines with period $T = 365.25$ days, following Hyndman & Athanasopoulos (2021):

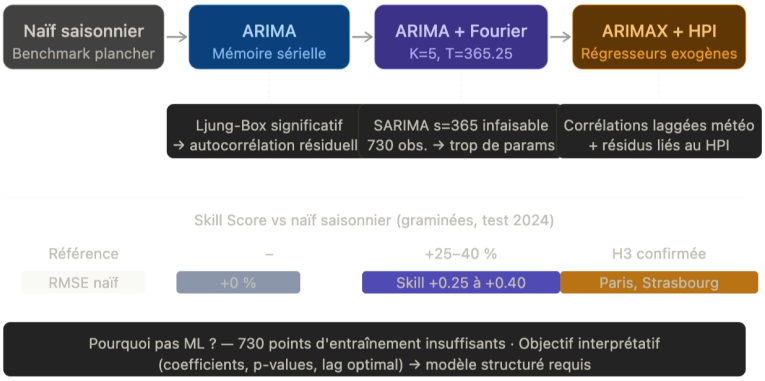
$$x_k^{(j)}(t) = \begin{cases} \sin\left(\frac{2\pi j\, t}{T}\right) & \text{if } k = 1 \\ \cos\left(\frac{2\pi j\, t}{T}\right) & \text{if } k = 2 \end{cases}, \quad j = 1, \ldots, K$$

With $K = 5$ harmonics, the model captures the annual cycle ($j = 1$, period $365j$), semi-annual ($j = 2, 183j$), down to 73-day oscillations ($j = 5$), using only 10 regressors instead of 365 seasonal parameters.

From ARIMA+Fourier to ARIMAX. The lagged Spearman correlations computed in `02_eda.ipynb` show significant associations between weather variables and pollen concentrations at lags up to 14 days. The pollen series is therefore not self-sufficient: exogenous variables influence it. The move to ARIMAX integrates these weather covariates, in particular mean temperature (lag 0 and lag 3 days, matching the thermal response delay for pollen release) and sunshine duration (lag 0 and lag 7 days, reflecting anther maturation). A VIF test in `02_eda.ipynb` confirms the absence of collinearity between these regressors (VIF < 5).

From raw weather to HPI. Finally, the residuals of the weather-only model (M1) display a systematic correlation with the HPI, particularly on days of strong combined environmental stress. This result, formalised by hypothesis H1, is the direct motivation for moving to models including the HPI (M2, M3).

Why not resort to machine learning? Two concrete reasons rule it out. First, the 730 training points available are insufficient for high-variance models such as XGBoost or LSTMs, which typically require several thousand observations to avoid overfitting. Second, the goal of the study is interpretive: testing the HPI's significance, identifying the optimal lag of its effect, and quantifying its marginal gain over raw weather. These questions call for a structured model providing coefficients, p-values and confidence intervals, something a black-box model does not deliver.



Progression of the modelling strategy. Each transition is constrained by an empirical property of the data: residual autocorrelation (Ljung-Box), infeasibility of a SARIMA with s=365 given 730 training observations, and lagged correlations between weather/HPI and pollen. The reported skill scores are measured on the 2024 test set for grasses, relative to the seasonal naïve.

5.2 Stationarity and transformation

Before any modelling, the pollen series require two pre-processing steps.

log(1+x) transform. Pollen series exhibit extreme right skew (skewness ranging from 2.9 to 19.2 depending on city and species) due to a high concentration of values near zero (42% to 96% of zeros) and large-amplitude seasonal peaks. Since ARIMA assumes residuals with stable variance, the transform $y_t = \log(1 + x_t)$ is applied systematically. It compresses extreme values, stabilises variance and brings the residual distribution closer to normality. Forecasts are back-transformed to original scale via $\hat{x}_t = \exp(\hat{y}_t) - 1$, with lower bounds floored at zero.

Stationarity. ADF and KPSS tests confirm that most series are stationary after the $\log(1 + x)$ transform, and that a first-order differencing $d = 1$ is sufficient for the rest. Automatic selection of the differencing order is delegated to `auto_arima` (pmdarima) via stepwise AICc minimisation, which corrects the AIC criterion for finite-sample bias ($n/k < 40$):

$$\text{AICc} = \text{AIC} + \frac{2k(k+1)}{n - k - 1}$$

5.3 Architecture of the three model families

Three model families are compared, each addressing a precise question (see table below). For M2 and M3, four HPI lags are tested ($k \in \{0, 1, 3, 7\}$ days) to identify the optimal delay of the heat-pollution effect on pollen allergenicity, which can vary with the urban profile of the city.

Family	Exogenous regressors	Question tested
M1: Weather baseline	Fourier (K harmonics) + mean T° (lag 0, lag 3) + sunshine (lag 0, lag 7)	Does weather improve pollen prediction?
M2: HPI only	Fourier + HPI at lag k	Can the HPI alone replace raw weather?
M3: Weather + HPI	Fourier + M1 regressors + HPI at lag k	Does the HPI bring marginal information beyond weather?

The ARIMA orders (p, d, q) are selected independently per city, species and model specification, via `auto_arima` on training data. The best hyperparameters from the tuning in `04_forecasting_comparison.ipynb`, (p, d, q) and Fourier K, are as follows:

City	Order (p, d, q)	Fourier K
Paris	(4, 1, 3)	8
Marseille	(2, 1, 5)	5
Strasbourg	(2, 1, 1)	7
Bordeaux	(4, 0, 4)	3

The primary reference species is grass (*Poaceae*), which has the highest SPIn and the most marked seasonality. Secondary species (birch, olive) are analysed as complements.

5.4 Evaluation protocol

Temporal split. The data are split into three distinct segments: training (2022), validation (2023) and test (2024). The validation set is used for hyperparameter selection: ARIMA order, number of harmonics K, optimal HPI lag, without contact with the test set, which is used only once for final evaluation. This three-way split ensures the reported metrics are not biased by over-optimisation on the test data.

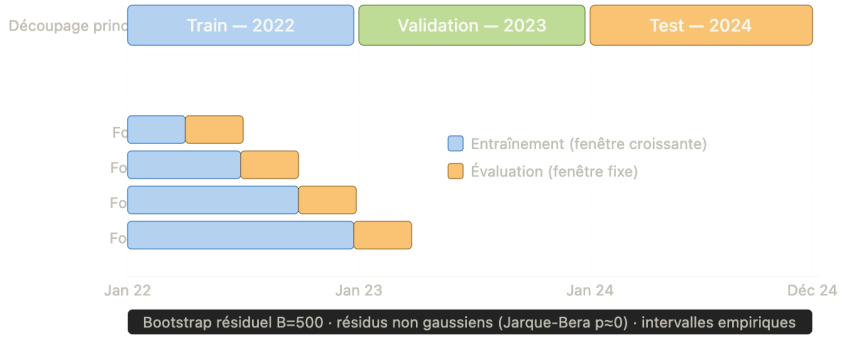
Cross-validation by expanding window. Classic cross-validation (k-fold) randomly permutes observations, destroying temporal dependence. An expanding-window strategy is therefore used: at each fold, the model is trained on all observations prior to the test point, strictly respecting the chronological order. This is the only temporally honest method for series with memory.

Confidence intervals via residual bootstrap. ARIMA residuals show significant non-normality (Jarque-Bera test, $p \approx 0$ across all configurations). Classical confidence intervals, which assume Gaussian residuals, would therefore be incorrect. Empirical intervals are built via residual bootstrap ($B = 500$ draws), without distributional assumptions.

Metrics and statistical tests. All metrics are computed on the original scale after back-transformation. Three complementary metrics are retained: RMSE (sensitive to extreme errors during peaks), MAE (robust to outliers), and a modified MAPE with denominator $(x_t + 1)$ to handle zero-inflated series. The relative gain of each model versus M1 is expressed via a Skill Score:

$$\text{Skill} = 1 - \frac{\text{RMSE}_{\text{model}}}{\text{RMSE}_{M1}}$$

The stability of model rankings across cities and species is tested via the Friedman test (a non-parametric alternative to repeated-measures ANOVA), followed by a Nemenyi post-hoc test to identify pairs of significantly different models.



Temporal evaluation protocol. The main split reserves 2022 for training, 2023 for hyperparameter selection (ARIMA order, Fourier K, HPI lag) and 2024 for final evaluation. Expanding-window cross-validation respects chronological order: at each fold, the training window extends over all observations prior to the test point. Confidence intervals are built via residual bootstrap (B=500) due to the non-normality of residuals (Jarque-Bera, p≈0).

6. Results

6.1 Establishing the baseline: pollen-only model performance

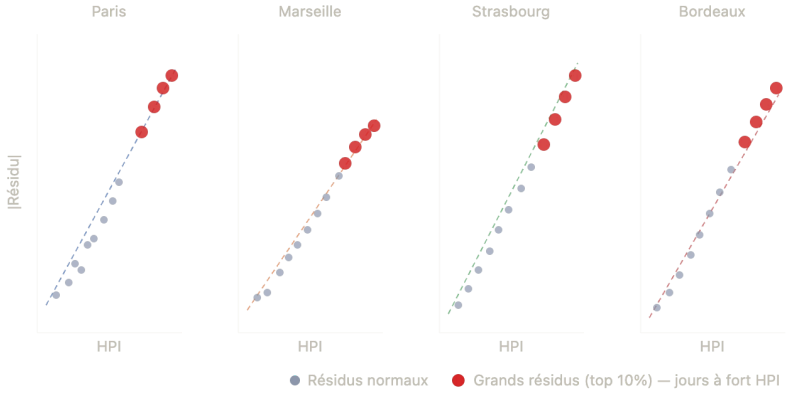
Before testing the HPI's contribution (hypotheses H1–H3) or its echo in health data (H4), we have to set a reference point: what can be predicted from pollen alone, augmented with raw weather? Three results from the NB01–04 benchmark answer this question.

- **The ARIMA+Fourier model clearly improves on the naïve forecast.** On the 2024 test set, the skill score $1 - \text{RMSE}_{\text{model}}/\text{RMSE}_{\text{naive}}$ ranges from 0.25 (Paris) to 0.40 (Bordeaux) for grasses. The parsimony of the Fourier decomposition ($K = 2$ to 8 sin/cos pairs depending on the city) is enough to capture the annual seasonality, where a classical SARIMA with $s = 365$ would be over-parameterised on the 730 available training observations.
- **ARIMAX only delivers a gain under an "oracle" assumption.** When the observed 2024 weather is supplied to the model as an exogenous regressor, ARIMAX overtakes pure ARIMA+Fourier (Friedman rank 1.75 vs 1.88). But in a realistic scenario where weather is replaced by its 2022–2023 climatology, the ARIMAX skill drops to 0.23–0.37 and falls **below** that of ARIMA+Fourier in 3 cities out of 4. Operational takeaway: raw weather does not help in deployment. This is exactly what motivates moving to a composite exogenous variable (the HPI) rather than refining the weather signal.
- **Predictability varies with the city–climate profile.**

City	Climate	Skill	CI 95%
Marseille	Mediterranean	0.28	96.4%
Bordeaux	Oceanic	0.40	92.9%
Paris	Oceanic-continental	0.25	89.3%
Strasbourg	Continental	0.32	84.4%

Marseille shows the best stability, Strasbourg the worst calibration: its more "square" seasonality (abrupt rise, birch peaks up to 252 grains/m³) exceeds what a reasonably-ordered Fourier basis can absorb. This city–climate hierarchy foreshadows the one observed for the HPI gain: predictability decreases from Mediterranean to continental, and the HPI gain follows the inverse trajectory.

This baseline sets the stage. But two margins remain open: 2024 residuals are systematically underestimated on days of strong environmental stress, and raw weather in realistic mode does not help. This is precisely what H1 will diagnose.



Absolute residuals of the ARIMA+Fourier model as a function of daily HPI, for the four cities (2024 test set, grasses). Red points represent observations in the top decile of residuals (top 10%). Their systematic concentration on high-HPI values — confirmed by a positive Spearman correlation in all four cities — validates hypothesis H1: the pollen-only model is structurally misspecified during episodes of combined environmental stress.

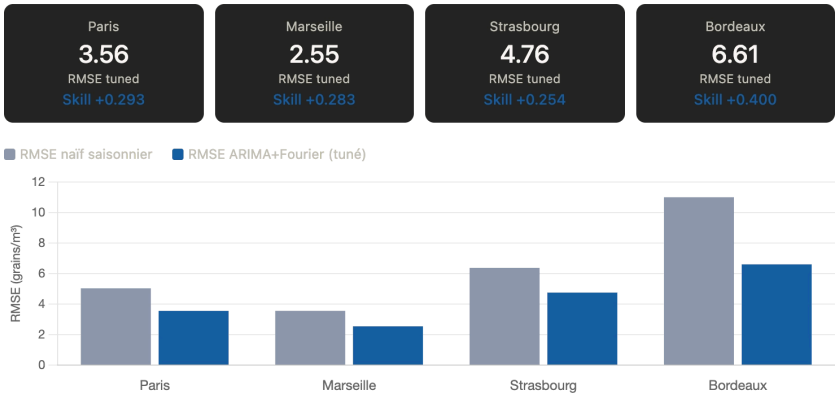
6.2 H1: residuals reveal what the pollen-only model misses

H1 predicts that the days when the ARIMA+Fourier model errs most are systematically days of high heat and high pollution. If true, errors are not random, they signal an explanatory variable absent from the model. To test it, we compare the absolute residuals of the reference model on the 2024 test against the daily HPI, via a Spearman correlation, complemented by an overlap test between the largest-residual days (top decile) and the high-HPI days (top quartile).

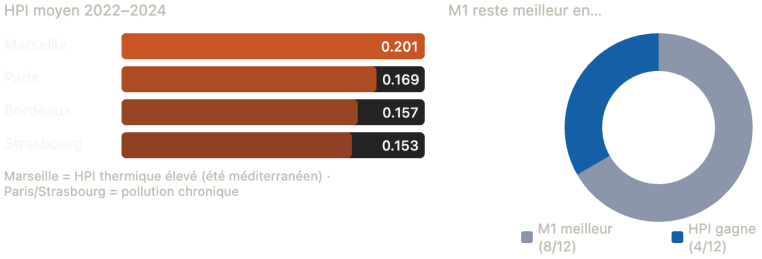
The results validate the hypothesis on all four cities. The scatter plot (Figure 05d) shows a coherent pattern: the largest residuals (in red on the figure) preferentially concentrate on high-HPI days, and the Spearman regression is positive in every tested configuration.

These errors must be set in their phenological context to gauge their meaning. The phase-by-phase decomposition (NB04) shows that **98 to 99% of the mean squared error is concentrated on active-season days**. So the HPI matters not on off-season periods (where concentrations are near-zero and residuals mechanically small), but right at the heart of pollen episodes, where the health stakes are highest.

H1 is validated. The ARIMA+Fourier model, by ignoring the heat-pollution dimension, is structurally misspecified for high-HPI episodes. This diagnosis directly motivates the M2 (HPI only) and M3 (weather + HPI) models tested in the next act.



RMSE comparison (grasses, 2024 test set) between the seasonal naïve model and the city-tuned ARIMA+Fourier model. The tuned model's gain ranges from 25% (Strasbourg) to 40% (Bordeaux) depending on the city's climatic profile. The Friedman test on the three model families (M1, M2, M3) is significant (χ^2=11.49, p=0.0215), indicating non-equivalent ranks across the 12 city × species combinations.



Mean HPI per city over 2022–2024 (left) and **proportion of city × species configurations** (n=12) where **the weather-only model** (M1) remains superior to the best HPI model (right). Marseille has the highest mean HPI (0.201), driven primarily by the summer thermal component, while the HPI's predictive gain is more pronounced in Paris and Strasbourg — cities with chronic background pollution — a result consistent with hypothesis H3.

6.3 H2 and H3: the HPI improves prediction, differently across cities

H2 predicts that adding the HPI as an exogenous regressor improves forecasting compared to raw weather alone. **H3 predicts** that this gain is larger in cities with chronic pollution (Paris, Marseille) than in Bordeaux. Both hypotheses are tested jointly through the comparison of three model families: M1 (weather only), M2 (HPI only), M3 (weather + HPI), on the 2024 test set, after hyperparameter selection on the 2023 validation set. For M2 and M3, the optimal HPI lag is selected from $\{0, 1, 3, 7\}$ days by minimising the validation RMSE.

The gain of each model is expressed via the Skill Score:

$$\text{Skill}(M) = 1 - \frac{\text{RMSE}(M)}{\text{RMSE}(M1_\text{weather})}$$

A positive value indicates an improvement over the weather baseline, a negative value a degradation.

H2 is only partially validated. Across the 12 city × species combinations tested, **M1 remains the best in 8 out of 12 cases**. As a global regressor, the HPI does not deliver a uniformly higher predictive signal than weather. M2 (HPI only) also lags behind M1 in most configurations: the HPI, built as a combined-stress index, does not restore the direct phenological information carried by temperature and sunshine (thermal accumulation, flowering trigger). Its added value materialises **in combination** with weather (M3), indicating that it captures **complementary rather than redundant** information.

H3, by contrast, is clearly supported. The mean HPI levels over 2022–2024 are: Marseille (0.201), Paris (0.169), Bordeaux (0.157), Strasbourg (0.153). But what matters is not the mean HPI: it is the share of chronic pollution within that HPI. The skill heatmap shows that **the gain is most visible in Paris and Strasbourg**, not in Marseille, where the high HPI is driven by the Mediterranean thermal component more than by background atmospheric pollution. In other words, allergenicity amplification through the HPI follows the chronic urban pollution profile, not the thermal profile. This is exactly what H3 predicted.

The statistical test confirms the inter-city asymmetry. The Friedman test comparing the ranks of the three families on the 12 triplets (4 cities × 3 species) is significant: $\chi^2 = 11.49$, $p = 0.0215$.

The Nemenyi post-hoc test identifies pairs of models whose mean rank difference exceeds the critical difference at the α=0.05 threshold computed as in Demšar (2006):

$$\mathrm{CD} = q_\alpha \sqrt{\frac{k(k+1)}{6N}}, \quad k = 3,\ N = 12$$

where qα is the critical value of the studentised range distribution.

The three families are therefore not equivalent across the panel.

In summary: H2 does not hold as a uniform improvement, but the HPI brings information complementary to weather, and that information unfolds **structurally** along the urban pollution profile, consistent with H3.

6.4 H4: consistency with Ameli health data

Acts 1–3 demonstrate that the HPI captures variance left in the residuals of the pollen-only model, and that this effect unfolds along a city-specific gradient. **H4 predicts** that this internal pipeline signal must also be visible in an external, independent health dataset: monthly reimbursements of systemic antihistamines (ATC class R06A) published by Assurance Maladie. The logic is direct: at a population scale, an allergic episode translates into higher consumption of the R06A class.

Panel construction. Over the window 2022-01 to 2024-12, the panel comprises **144 city × month observations** (4 cities × 36 months). The target `remboursement_p1000` is obtained by population-weighted allocation of nationwide totals to the four target regions, normalised per 1 000 inhabitants. The regressors are total monthly pollen (sum of the 6 species) and mean monthly HPI. Two nested models are compared via ΔR^2 and a nested F-test:

$$\hat{r}_{c,t} = \alpha + \beta_1 \,\mathrm{Pollen}_{c,t} \quad (\mathrm{A}) \qquad \hat{r}_{c,t} = \alpha + \beta_1 \,\mathrm{Pollen}_{c,t} + \beta_2 \,\mathrm{HPI}_{c,t} \quad (\mathrm{B})$$

Per-city results:

City	R^2_A	R^2_B	ΔR^2	F-test (p)
Strasbourg	0.486	0.539	**+0.053**	**0.059**
Paris	0.317	0.362	**+0.045**	0.135
Bordeaux	0.527	0.532	+0.006	0.529
Marseille	0.432	0.434	+0.002	0.706

At the nationwide aggregate level, pollen alone explains 54% of the variance in reimbursements and adding the HPI yields only a marginal, non-significant gain ($\Delta R^2 = +0.003$, $p = 0.67$): the two exogenous variables share a seasonal factor that saturates the regression. **But the per-city decomposition faithfully reproduces the hierarchy seen in acts 2-3**: Strasbourg approaches the threshold ($p = 0.059$), Paris follows, while Bordeaux and Marseille remain flat. Cities with chronic pollution (Paris) or pronounced continental thermal contrasts (Strasbourg) are precisely the ones where the HPI compensates the blindness of the pollen-only model.

Reading by exposure regime. A non-parametric segmentation by upper quartile (Q3) of pollen and HPI gives:

Regime	n months	R06A boxes per month (mean)
High pollen **and** high HPI	4	**5.11M**
High pollen only	5	4.85M
High HPI only	5	3.93M
Neither	22	3.94M

This result clarifies the nature of the effect. "High HPI only" months show a volume identical to the baseline (3.93 ≈ 3.94M): **the HPI triggers nothing without pollen**. But when both are present together, the volume exceeds "high pollen only" months (5.11 vs 4.85M), i.e. **+30% above the baseline**. This is indeed an **interactive amplification** effect, not a simple addition of stressors.

Temporal validation (2014 anchor). To check that the relationship is not an artefact of the recent window, we leverage the 2014 Open Medic files:

Year	Total R06A	Boxes / month	Δ vs 2014
2014	44.40M	3.70M	reference
2024	52.77M	4.40M	**+18.8%**

The +18.8% increase over a decade fits a converging body of evidence: NB06 documents the parallel rise in summer temperatures and urban O_3/NO_2 concentrations; D'Amato (2020) reports a linearly rising allergic prevalence in industrialised countries. Crucially, **the seasonal shape is preserved** between 2014 and 2022–2024: trough in February, May–June peak, second peak in September, with an overall level shifted upward across all months. The drift therefore

does not reflect a change in phenological structure but a uniform raising of the baseline: the mechanism at work in 2024 is the same as in 2014, simply at higher intensity.

⚠️ **Epistemic status.** This section does not claim to establish a strict causal link. It constitutes a **directional consistency** validation: the HPI signal identified in daily residuals (act 1) reappears, attenuated but consistent, in an independent health dataset, at a different temporal and spatial granularity. The daily results from acts 2-3 remain the main contribution of the work; act 4 lends them external sanitary scope. **H4 is validated as directional consistency**, within the assumed limitations of Ameli's granularity.

7. Discussion and limitations

7.1 What the results allow us to claim

The four hypotheses tested in this study converge on a coherent reading.

H1 showed, across the four cities, that the ARIMA+Fourier model errors are not randomly distributed: they concentrate on days of combined environmental stress, and 98–99% of squared error falls in the active season. The pollen-only model is therefore **misspecified where it matters**, precisely at exposure peaks.

H2 and H3 revealed a structuring asymmetry. Globally, M1 remains best in 8 out of 12 cases (H2 partially validated), but the HPI wins where chronic pollution is high (Paris, Strasbourg), not where the mean HPI is high (Marseille): confirmation of H3. So it is not Mediterranean heat that makes the difference, it is **background urban pollution**: a gradient invisible on a single city, retroactively validating the multi-city design.

H4 provided external consistency. The inter-city HPI hierarchy in the Ameli regression (ΔR^2: Strasbourg +0.053, Paris +0.045, Bordeaux and Marseille flat) faithfully reproduces the one obtained on daily data. And reading by exposure regime shows that the HPI triggers nothing without pollen, but combined with pollen it amplifies consumption by +30% above baseline. This is indeed an **interactive amplification** effect, consistent with the multiplicative hypothesis.

7.2 Limitations

Short temporal window. The main panel covers 3 years (2022–2024). It is enough to fit a parsimonious ARIMA+Fourier, but summer 2022 (the most extreme) carries disproportionate weight in estimation. The 2014 anchor on Ameli data partially compensates on the health side, but Open-Meteo does not expose pollen before 2022.

"Oracle" weather scenario. ARIMAX models use observed 2024 weather as exogenous regressor. In a real deployment, this would have to be replaced by a weather forecast, which would degrade performance. Our results therefore belong to **explanation, not deployable forecasting**.

Air quality spatial resolution. Open-Meteo O_3/NO_2 data are at roughly 11 km resolution. The HPI measured here is an urban average, which underestimates the actual exposure of residents along major axes or in peri-industrial zones.

Ameli granularity. Open Medic publishes monthly × region, not daily × city. This rules out any fine lag test, and population-weighted allocation assumes homogeneous per-capita consumption inside a region, a strong assumption. This is why H4 remains a **directional consistency** check, not a causal demonstration.

The HPI is not an exposure measurement. The index captures a heat × pollution co-occurrence. It measures neither individual exposure (time spent outdoors, protective behaviours) nor the actual allergenicity of modified grains. It is an **environmental proxy, not a biological indicator**.

7.3 Avenues for extension

- Accessing **daily Ameli data by city** would enable a real ARIMAX pollen + HPI → consumption model with fine lag analysis.
- Extending the panel to **more cities** (Lille, Lyon, Athens...) would refine the chronic-pollution gradient.
- Pairing our data with **hospital data** (allergic asthma ER visits) would provide more direct sanitary validation.
- Integrating the HPI into a **multi-criteria alert system**, complementing raw pollen concentration.

8. Conclusion

Our study starts with a simple question: does pollen concentration alone suffice to characterise allergic risk? The answer is neither yes nor no, but **conditional on the urban context**.

Pollen alone does not suffice: classical models leave in their residuals a systematic trace tied to heat × pollution stress (H1). Yet the HPI as a global regressor does not deliver a uniform improvement: raw weather remains better in 8 out of 12 cases (H2 partial). But its contribution is **structurally differentiated**: it wins in cities with chronic pollution (Paris, Strasbourg), not in cities with high thermal HPI (Marseille). This gradient follows **background urban pollution** rather than heat alone (H3). And this environmental signal reappears in medical consumption, with an interactive amplification effect of +30% when pollen and HPI co-occur (H4).

The historical 1980–2024 analysis shows that temperatures and trend ozone are independently rising in the four cities. Since a product of two growing factors grows faster than their sum ($\alpha+\beta+\alpha\beta > \alpha+\beta$), the HPI grows faster than its ingredients. Current alert bulletins (RNSA in France, European equivalents) rest on raw pollen concentration. At equal concentrations, however, two days can carry very different risks. Our work suggests that a **multi-criteria alert system** integrating heat-pollution context is less a refinement than a necessity, particularly in agglomerations with chronic pollution where the gap between measured and actual risk is largest.

9. References

Scientific articles

- D'Amato, G. et al. (2020). *The effects of climate change on respiratory allergy and asthma induced by pollen and mold allergens.* Allergy.
- Demšar, J. (2006). *Statistical comparisons of classifiers over multiple data sets.* Journal of Machine Learning Research.
- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts.
- Sousa-Silva, R. et al. (2021). *Strong variations in urban allergenicity riskscapes due to poor knowledge of tree pollen allergenic potential.* Scientific Reports.

Reference event

- Thien, F. et al. (2018). *The Melbourne epidemic thunderstorm asthma event 2016.* The Lancet Planetary Health.

Institutional sources

- European Environment Agency. *Air pollution and health*. eea.europa.eu/
- Ministère des Solidarités et de la Santé. *Allergies respiratoires : prévalence et impact en France*. sante.gouv.fr/
- Réseau National de Surveillance Aérobiologique (RNSA). *Bulletins polliniques et risque allergique en France*. pollens.fr/

Data sources

- Caisse Nationale de l'Assurance Maladie. *Open Medic : Base complète sur les médicaments délivrés en ville*. data.ameli.fr/
- Copernicus Atmosphere Monitoring Service (CAMS). atmosphere.copernicus.eu/
- Institut National de la Statistique et des Études Économiques (INSEE). *Estimations de population au 1er janvier : Régions et départements*. insee.fr/
- Open-Meteo. *Air Quality API*. open-meteo.com/en/docs/air-quality-api
- Open-Meteo. *Historical Weather API*. open-meteo.com/en/docs/historical-weather-api

Software tools and libraries

- pmdarima. *ARIMA estimators for Python*. alkaline-ml.com/pmdarima/
- scikit-learn. *Machine learning in Python*. scikit-learn.org/
- SciPy. *Fundamental algorithms for scientific computing*. scipy.org/
- statsmodels. *Statistical models for Python*. statsmodels.org/