Diurnal, Seasonal and Spatial Patterns of Weather Influences on PM2.5 Concentrations in Poland: An Explainable Machine Learning Approach

Tetiana Vovk* & Maciej Kryza

Faculty of Earth Sciences and Environmental Management, University of Wrocław

* Contacting author: tetiana.vovk@uwr.edu.pl

Uniwersytet Wrocławski

INTRODUCTION

- The relationship between meteorological conditions and $PM_{2.5}$ concentrations is complex, involving nonlinear interactions and seasonal variability. Traditional statistical methods often fail to capture these dynamics fully.
- Recent advances in explainable machine learning, particularly the SHAP (SHapley Additive exPlanations) framework, offer a novel pathway to interpret model predictions and disentangle variable importance in environmental systems (Lundberg & Lee, 2017; Molnar, 2022).
- In this study, we analyze a decade of hourly $PM_{2.5}$ and meteorological data (2014–2023) from 11 cities across Poland, applying SHAP to reveal **how weather features modulate air pollution under varying temporal and spatial contexts.** The objective is to move beyond simple correlation, and instead quantify **how**, **when**, and **where** different meteorological drivers affect $PM_{2.5}$ formation and dispersion.
- This approach allows us to identify both consistent physical relationships and emergent, context-specific behaviors consistent with atmospheric chemistry and boundary-layer dynamics (Petäjä et al. 2016).

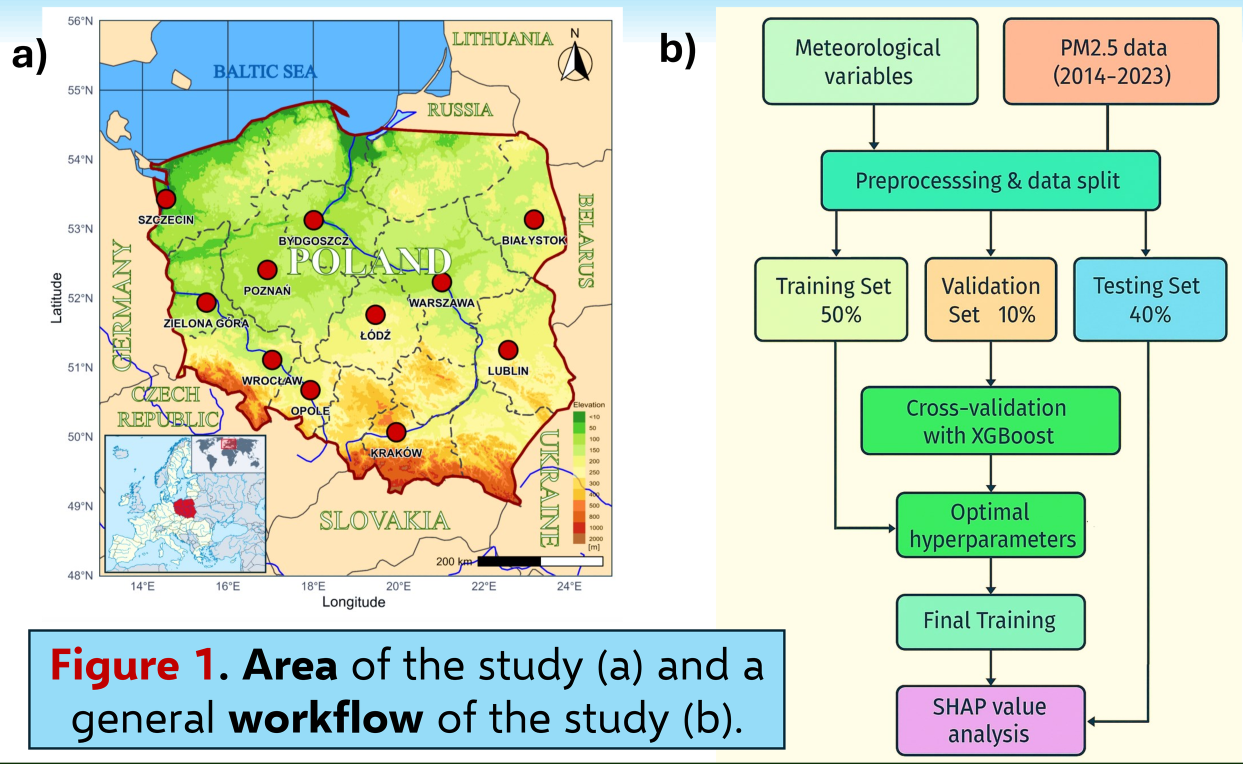

Figure 1. **Area** of the study (a) and a general **workflow** of the study (b).

DATA & METHODS

- ✓ We trained a gradient boosting model (**XGBoost**) to predict **hourly $PM_{2.5}$ concentrations** based on meteorological inputs from the **WRF** (Weather Research and Forecasting) model.
- ✓ For each forecast hour, the model used both **instantaneous values** and **24-hour rolling means** of: *precipitation, 2-m temperature, shortwave radiation, wind speed, wind direction, planetary boundary layer height, relative humidity, sea-level pressure* **as predictors**.
- ✓ **Hourly $PM_{2.5}$ data (2014–2023)** used as the dependent variable came from CEIP (Chief Inspectorate of Environmental Protection, Poland) at 11 cities (Fig. 1,a). Each location included 1–5 stations; data were spatially averaged per city.
- ✓ **10% of data** were used for **temporal cross-validation** to tune six hyperparameters: *max_depth, sample_size, Mtry, min_N, learn_rate, Ntree*. **50% of data** were used for training, **40%** – for testing. The full workflow summary is shown in Fig 1, b.
- ✓ The final model achieved mean absolute error (MAE) of **5.35 $\mu g \cdot m^{-3}$** and R^2 of **0.79**.
- ✓ Using the *caret* package in R, we applied SHAP (**SHapley Additive exPlanations**) to the test predictions. SHAP values decompose each prediction into **individual variable contributions**. This method calculates feature importance by comparing model predictions with and without each feature across *all possible subsets of features*. For each subset S, two models are trained: one including the feature i and one excluding it. **The prediction difference shows the feature's impact,** and averaging these differences over all subsets gives the Shapley value as the final feature attribution.

➤ **Positive SHAP → variable increases $PM_{2.5}$** ➤ **Negative SHAP → variable reduces $PM_{2.5}$**

RESULTS & DISCUSSION

DECADAL TRENDS

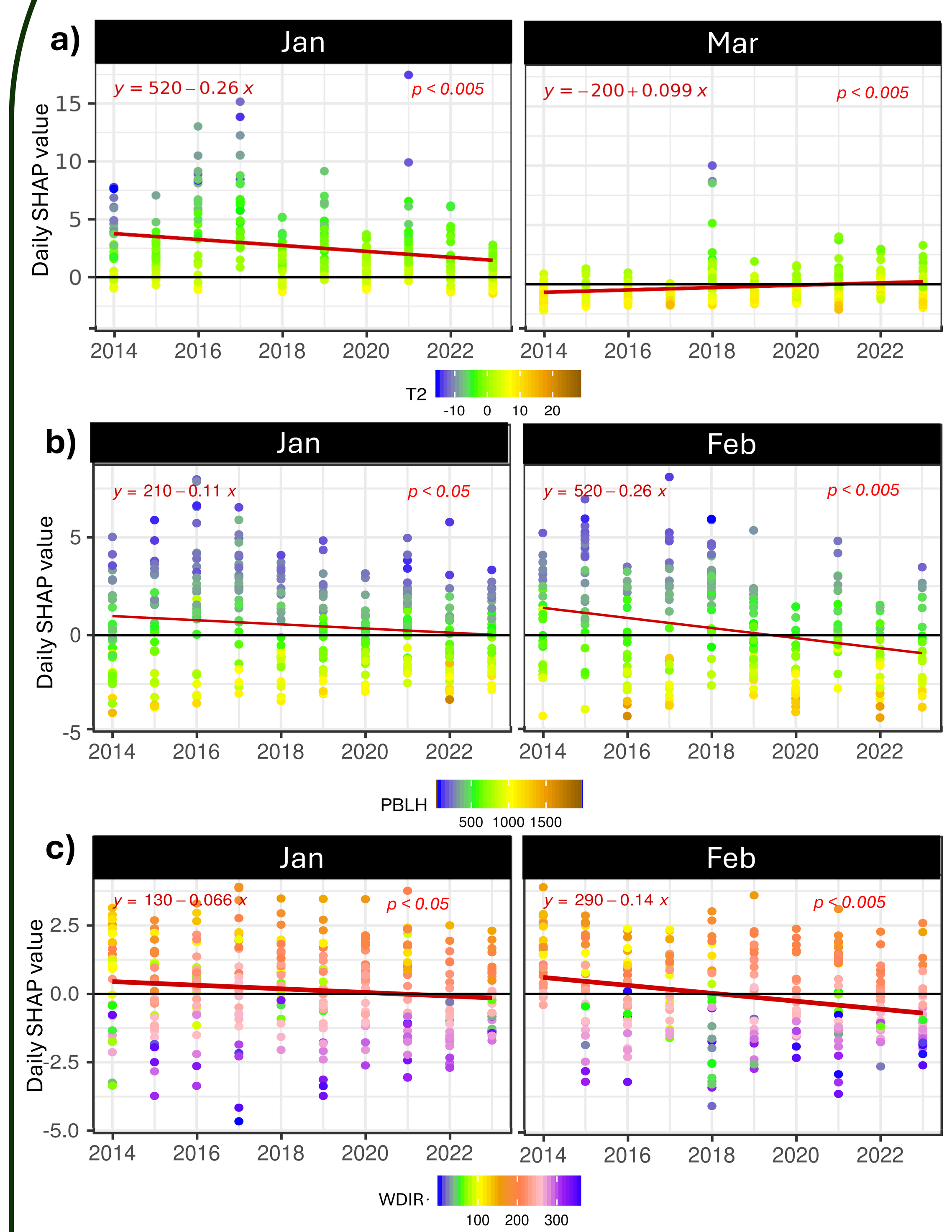

Figure 2. **Decadal trends** of **temperature** (a), **PBLH** (b), and **wind direction** (c) influences on PM2.5

<u>**Temperature influence**</u> in **Jan** has significantly declined over the decade, suggesting that warmer conditions are reducing cold-related $PM_{2.5}$ build-up. For months from **Mar to Aug**, significant **positive trends** were found instead, indicating the increasing positive temperature impact during spring and summer.

<u>**PBLH influence**</u> in **Jan and Feb** shows a declining or negative trend, indicating a growing role of atmospheric mixing and more variable synoptic conditions in mid-winter. Similar significant trends were found for **Oct and Nov.**

<u>**Wind direction influence**</u> in **Jan and Feb** is shifting toward cleaner patterns. Negative trend in wind direction implies more frequent NW/W/SW ventilation events, replacing stagnant continental easterly patterns.

DIURNAL AND SEASONAL PATTERNS

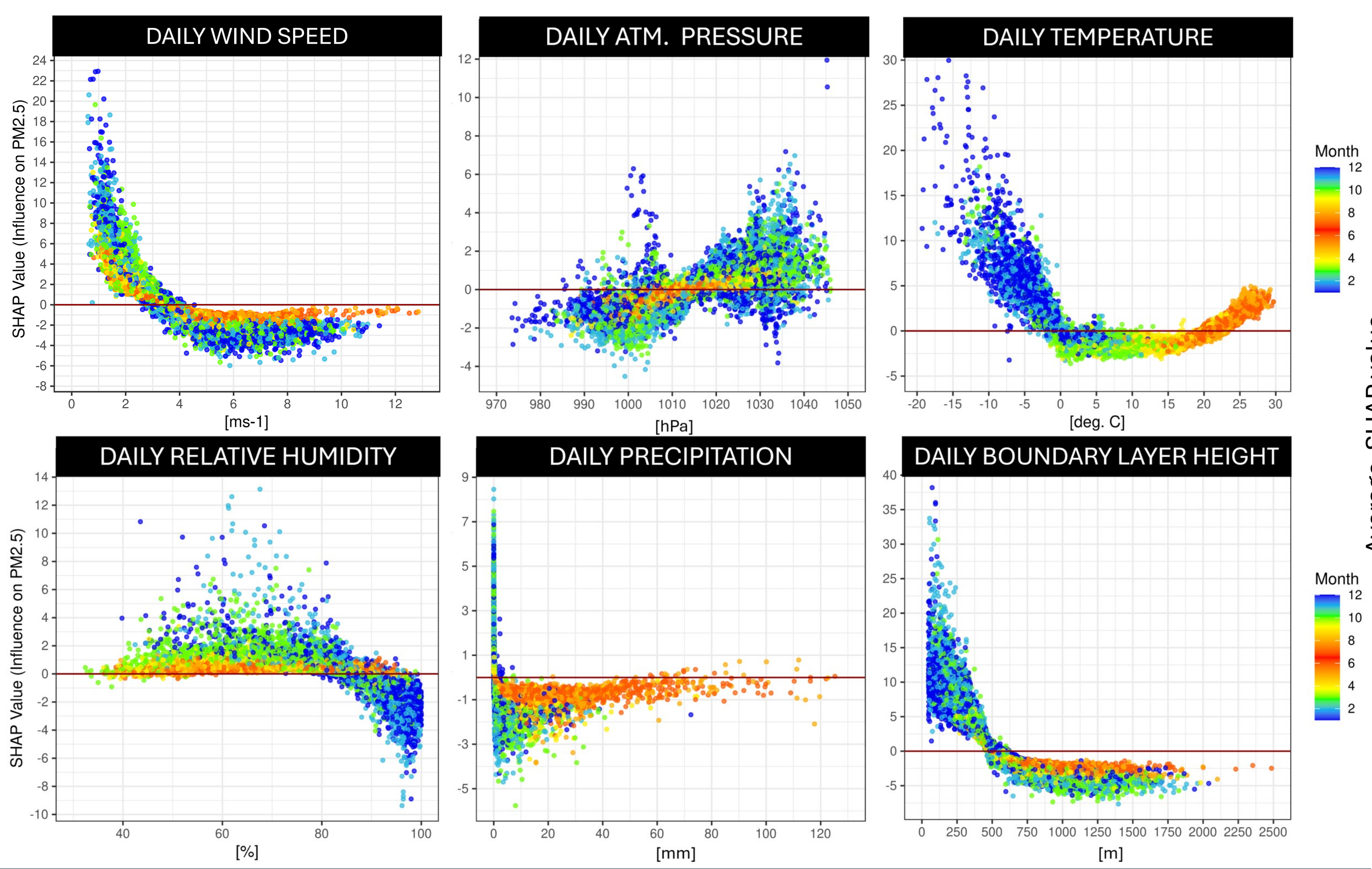

Figure 3. **Seasonal pattterns** of daily **wind speed, relative humidity, atmospheric pressure, precipitation, temperature** and **boundary layer height** (PBLH) influence on PM2.5.

<u>**Wind Speed**</u>

➤ The strongest positive effect on $PM_{2.5}$ occurs at wind speeds < 3 m/s across all seasons. At wind speeds > 4 m/s, the impact becomes negative. Effects are more pronounced in colder seasons, and weaker in summer, even at higher speeds.

<u>**Relative Humidity**</u>

➤ Daily RH shows peak influence on $PM_{2.5}$ (positive SHAP) at mid-range (55–70%) in transitional seasons. Very high daily RH (>90%) in cold seasons flips SHAP negative.

<u>**Atmospheric Pressure**</u>

➤ High pressure (>1020 hPa) aligns with higher $PM_{2.5}$ under stagnant anticyclonic conditions; low preassure is assosiated with lower $PM_{2.5}$, especially in transtional months.

<u>**Precipitation**</u>

➤ In cold seasons, even moderate rain reduces $PM_{2.5}$ (negative SHAP). In summer, this effect weakens—possibly due to much shorter but more intensive precipitation events and already low pollution.

<u>**Temperature**</u>

➤ SHAP values rise sharply below 0°C (suggesting heating & inversions) in winter. In the 0 °C – 20 °C range, the effect on $PM_{2.5}$ is predominantly negative. SHAP turns positive again in summer motnhs – when daily temperature exceeds ~20°C.

<u>**Boundary Layer Height (PBLH)**</u>

➤ PBLH below 500 m is linked to strong positive SHAP. Higher PBL flattens SHAP and turns it negative, showing cleaning effects of convective mixing, most pronounced in transitional months.

<u>**Wind Direction**</u>

➤ N/NW/W winds show mostly negative SHAP, suggesting cleaning effects of polar and marine air masses. E–SE–S winds show mostly positive influence on $PM_{2.5}$. SW/NE show both positive and negative contributions. The influence is most pronounced in winter months and the weakest in summer.

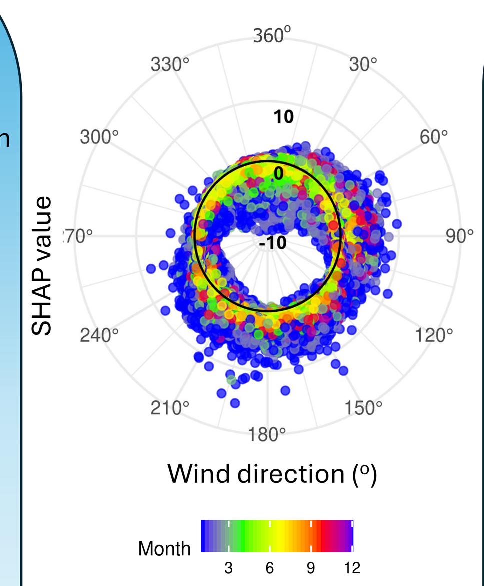

Figure 6. **Seasonal patterns of wind direction influence** on $PM_{2.5}$.

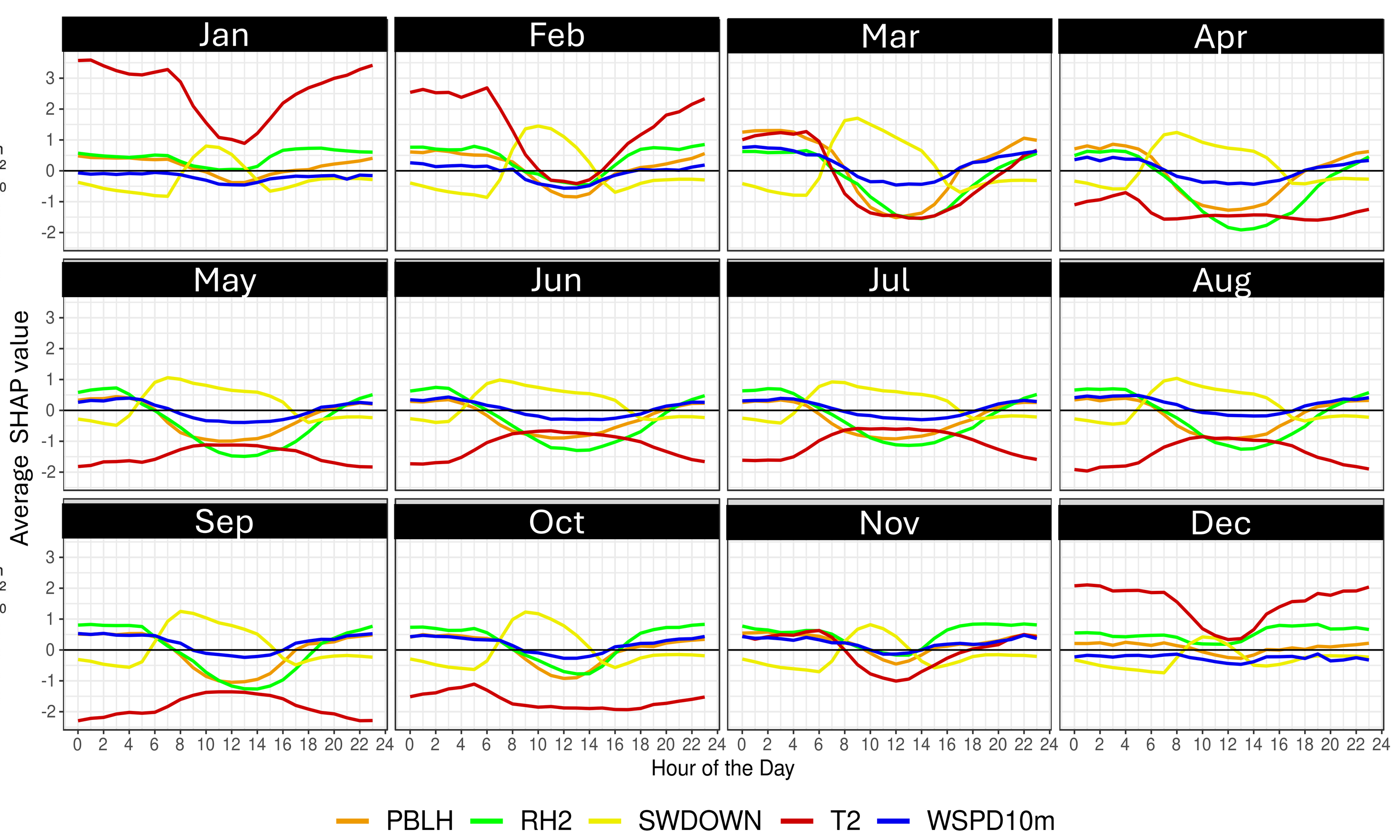

Figure 4. **Diurnal patterns** of **PBLH, relative humidity** (RH2), **shortwave radiation** (SWDOWN), **temperature** (T2) and **wind speed** (WSPD10m) **influence** on PM2.5 **by month**

<u>**Boundary Layer Height (PBLH)**</u>

➤ SHAP values are negative during the daytime — indicating that solar-driven PBLH rise reduces $PM_{2.5}$ by enhancing dispersion. The effect is most evident in early spring, associated with sharp daytime increases in PBLH. SHAP turns positive in early morning and evening hours, correlating with persistently shallow boundary layers. In winter, the diurnal variation of PBLH influence becomes less pronounced.

<u>**Relative Humidity (RH2)**</u>

➤ Generally positive SHAP values overnight and in the early morning, especially during cooler months. These turn more neutral or negative by midday, with the strongest variation seen in spring. The overall pattern resembles that of PBLH.

<u>**Shortwave Radiation (SWDOWN)**</u>

➤ Positive SHAP influence during the day, peaking from Feb to Apr — possibly indicating sunlight enhances photochemical PM2.5 formation (e.g., secondary aerosols). Nighttime influence is minimal, consistent with the lack of direct radiative effects.

<u>**Temperature (T2)**</u>

➤ The diurnal pattern forms an upward parabola from Nov to Mar, with lower or negative SHAP values during the day and high positive values at night — suggesting winter nights are linked to stable, polluted conditions (thermal inversions + combustion emissions). In summer, the trend reverses into a downward parabola: daytime may coincide with increased emissions, secondary aerosol formation, and resuspension, while nighttime tends to show reduced emissions.

<u>**Wind Speed (WSPD10m)**</u>

➤ Slightly negative SHAP values in the afternoon, especially in spring and autumn, indicating modest dispersion effects. Mostly positive overnight and in early morning hours — a pattern similar to PBLH and RH.

SPATIAL PATTERNS

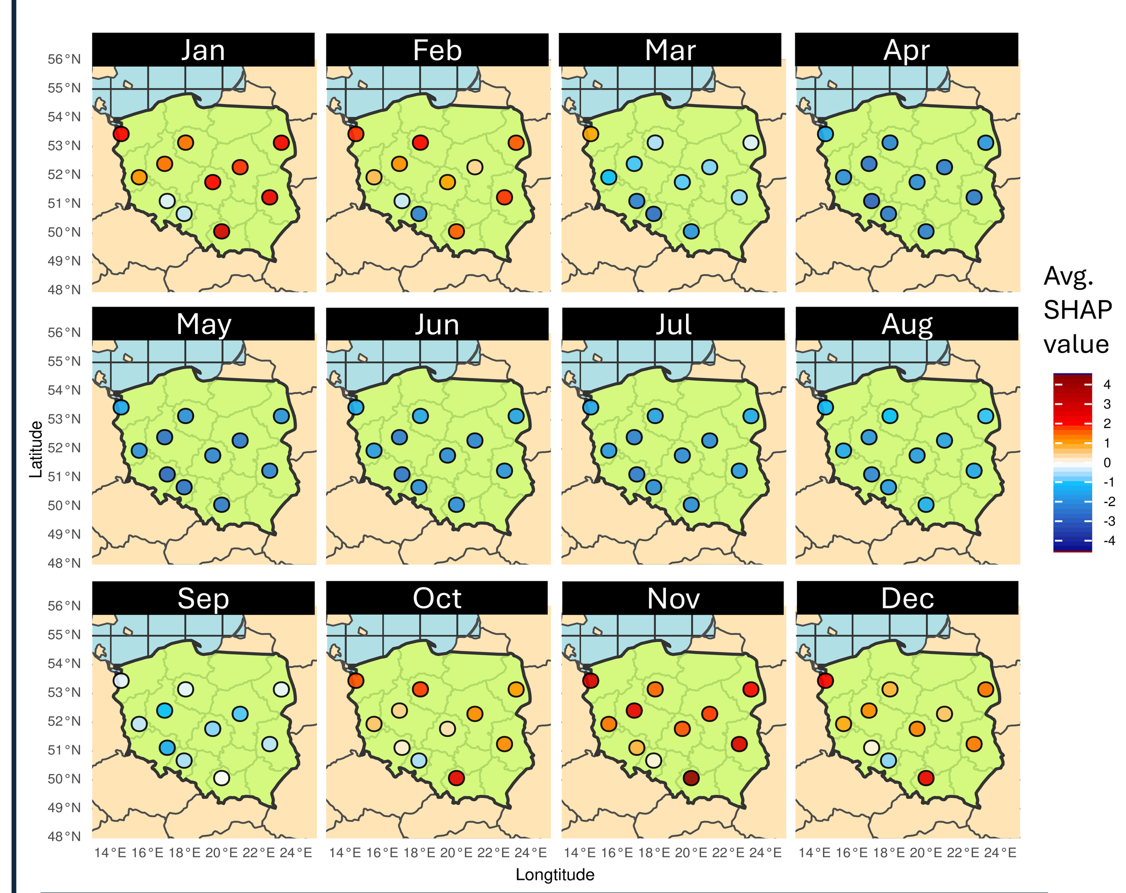

Figure 5. **Spatial patterns** of **PBLH influence** on PM2.5 shown **by month**

In autumn (since October): SHAP turns positive, which coincides with more often nighttime inversions and emissions rise. **In spring: SHAP turns negative** quickly, presumably, due to rapid increase in convective mixing. PBLH starts acting clearly as a dispersion driver from **March** onward.

Szczecin is an outlier with delayed response, showing positive SHAP in March. Likely due to coastal marine influence and shallower boundary layers, which make PBLH less effective for pollution reduction until later in the season.

Wrocław & Opole in south-west Poland show mostly negative SHAP, **even in winter**. These could be related to more frequent mixing-enhancing events and higher average PBLH compared to other locations during the decade.

CONCLUSIONS

- ✓ Explainable machine learning reveals **consistent physical rules** in how weather influences $PM_{2.5}$, highlighting the method's strength for interpreting air quality models.
- ✓ $PM_{2.5}$ responses to meteorology **are highly scale- and context-dependent**, with clear seasonal, diurnal, and spatial modulations.
- ✓ SHAP analysis captured **both expected patterns and subtle regime shifts**, such as later onset of dispersion effects in some cities or temperature-related $PM_{2.5}$ rise in warm conditions.
- ✓ Observed decadal trends suggested **evolving pollutant sensitivity**, likely linked to **changing emissions and climate factors**.
- ✓ Overall, the approach offers a **valuable framework to track, explain, and anticipate air quality dynamics** in a changing environment.

REFERENCES

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems, 30.
Molnar, C. (2022). Interpretable Machine Learning. Leanpub.
Petäjä, T., et al. (2016). Enhanced air pollution via aerosol-boundary-layer feedback in China. Scientific Reports, 6, 18998.

ACKNOWLEDGMENTS

This study was supported by Polish National Science Center, grant no *UMO-2021/43/B/ST10/01189.*