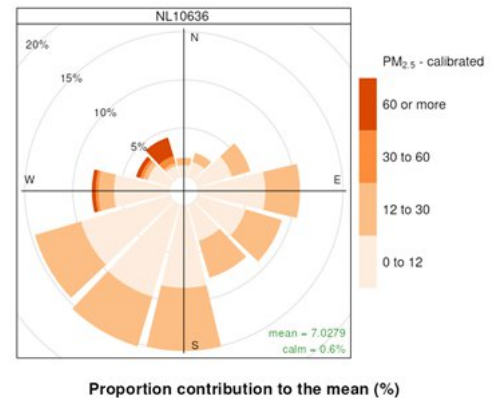


Challenge 2: Discover Pollution Hotspots

Challenge Description

Hollandse Luchten collects large amounts of sensor data on air quality. Residents often question the source of pollution they experience. Organizations such as the RIVM and GGD often use wind roses to detect these local sources. A wind rose shows the direction, speed and frequency of the wind captured at that sensor location. By combining wind and air pollution data you can create a pollution rose.



The plot displayed shows a weighted

concentration rose. It shows, for each wind direction, the share (in %) of that direction in the overall average particulate matter concentration. The average concentration for each wind direction is weighted by how often that wind direction occurs. When making a wind rose, using different choices for timesteps, such as daily, weekly or for example very specific time periods will have a big impact on the resulting pollution rose. It is also possible to plot multiple wind roses from different locations together to obtain an even clearer picture of where hotspots may be located.

Your challenge is to create a system that automatically analyses sequences of pollution roses over time to detect unusual hotspots and support source identification.

Problem Statement

Pollution roses are powerful tools for understanding local air quality patterns in relation to wind direction and speed. By visualizing these roses over extended time periods on a geographic map, you essentially create a "video" or time-series of frames showing how pollution patterns evolve.

Therefore, in this challenge you will focus on:

1. Using data science and AI techniques to automatically **detect** unusual patterns and potential pollution hotspots in sequences of pollution roses.
2. Developing a **scoring metric** that indicates how strong and unusual a hotspot is, and how certain the system is about the result.
3. **Locating** identified hotspots geographically and relating them to likely wind directions and possible source areas.

4. **Communicating** the **results** to residents in a clear and transparent manner, including why a location or direction is identified as a potential hotspot.
5. **Bonus**: indicating which additional data, analyses or follow-up actions are needed to confirm a potential pollution source.

Possible techniques

Teams may use statistical, machine-learning, deep learning and/or hybrid methods. For example, computer-vision techniques can be used to recognize spatial patterns in pollution roses, while unsupervised learning can help identify unusual patterns or deviations from normal conditions. Temporal analysis can be used to assess how pollution patterns develop over consecutive time periods.

A simple and well-validated approach may be more valuable than a complex model that is difficult to explain or verify. We expect teams to substantiate their technical choices and considerations using validation metrics, research or other evidence.

Which data is available?

During the hackathon we will supply you with several datasets for this challenge:

- Time-series data: hourly PM2.5 measurements from more than 100 Hollandse Luchten sensors, some ranging several years.
- Reference measurements: Official reference station data producing hourly measurements, to possibly compare the Hollandse Luchten sensors with.
- Metadata on sensors: information about the sensor location, type, measurement communities, and intended reference stations.
- Weather data from the Dutch KNMI, for example temperature and humidity.
- Some example pollution roses

Deliverables

Your submission should at least contain:

- 1) The Github repository with any used code and/or examples.
- 2) The presentation of your solution.

Every team is expected to present their solution to our jury.

Jury criteria

Will be added later

Use of tools and generative AI

AI (coding) assistants and vibe coding are allowed, as long as teams disclose clearly on where they did and did not use them. We encourage the use of such tools, but also address that teams should be able to explain and validate important parts of their solutions themselves.

Questions?

The full repository will be shared a week before the hackathon. If you have any questions, please e-mail them to janus.jansen@noord-holland.nl. These and spontaneous questions will be answered during our 'Online Q&A' on 6 November 15:00-16:00. Participants will receive the Q&A meeting link via email.