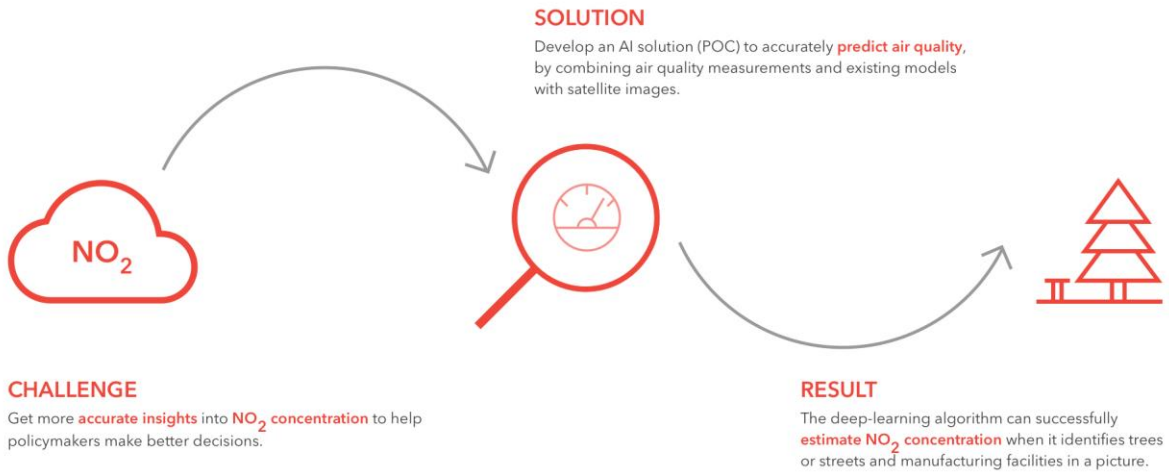
Predicting air quality using geospatial AI

With environmental concerns at an all-time high, the number of people concerned about the quality of the air we breathe is rising. Through geospatial AI, the European Environment Agency, delaware and Microsoft aim to get more accurate information about NO_2 concentration to help policymakers make better decisions.

Nitrogen dioxide is one of the main compounds produced by industrial activities – and a key indicator of polluted air. delaware experts worked with the European Environment Agency and Microsoft to develop a proof of concept AI solution capable of accurately predicting air quality at high resolution, by combining air quality measurements and existing models with satellite images.



1 Microsoft technologies drive the deep-learning solution

The team used the results of NO_2 modeling (2016 reanalysis) at a 7x7-kilometer scale together with 2,000 ground measurements to train a deep-learning model to estimate air quality at a 1x1-kilometer resolution. Our technology stack consisted of Azure Databricks with GPU support, Azure Maps, and the European Copernicus Atmosphere Monitoring Service (CAMS) dataset.

2 Augmenting air quality data with satellite images

CAMS delivers results from air quality modeling on a regular basis. Several EU countries gather a total of around 2,000 NO_2 measurements on the ground and report them to the European Environment Agency (EEA). This data was downloaded from the EEA's Air Quality e-Reporting system.

The most important source of information used to estimate annual NO_2 is CAMS modeling results. They are available for the whole of Europe and provide a good initial idea about how good or bad the air quality is.

"After gathering this data, we needed to understand the environment surrounding the sensors to offset satellite measurements," explains Wouter Labeeuw, data scientist at delaware. "We observed that more trees in an area will result in lower annual NO_2 measurements. In high-traffic areas, however, NO_2 increases as a byproduct of fuel combustion. Satellite imagery is an ideal solution because it offers a perfect view of the environment."

"Predicting Air Quality has become a lot more accurate if we not only take sensor (numeric) measurements, but combine it with maps and satellite images (visual) using deep learning. Instead of building a network of millions of sensors we can make almost street level air quality predictions by combining the right data sources, (artificial) intelligent algorithms, scalable compute and smart people. The smart people of delaware were essential in making this pilot a success!"

- Nathan Bijnens, Microsoft

The next step was to train an end-to-end deep-learning model on how to offset CAMS data using satellite imagery. The presence of many streets will mean that NO_2 measurements are predicted to be higher in a specific 1x1-kilometer zone compared to the 7x7-kilometer CAMS modeling results. "On its own, the deep-learning algorithm developed by the team successfully learned to reduce NO_2 estimation when it identifies trees in the picture, and increase NO_2 estimates in the presence of streets or manufacturing facilities," asserts Wouter.

3 Improving upon existing algorithms

The accuracy of the model's predictions was better than expected: the mean absolute error was cut in half and the root mean squared error was reduced by 40% compared to the currently used prediction algorithm. "There are still some working points for the method: it is possible to identify CAMS data in the model's predictions, so these need to be smoothed out in versions to come," Wouter adds.

"We would like to thank the European Environment Agency for giving us the opportunity to prove that our strategy works, and Microsoft for their support with the implementation of the solution on Azure," Wouter concludes.