



# BIG DATA

## FOR SOCIAL GOOD

### Case Study: Telefónica Brazil

## Predicting air pollution levels 24 to 48 hours in advance in São Paulo, Brazil



The GSMA represents the interests of mobile operators worldwide, uniting nearly 800 operators with more than 300 companies in the broader mobile ecosystem, including handset and device makers, software companies, equipment providers and internet companies, as well as organisations in adjacent industry sectors.

The Big Data for Social Good Initiative convenes public and private organisations to accelerate the mobile industry's impact against the UN Sustainable Development Goals (SDGs).

Infectious diseases, pollution, earthquakes, floods and other disasters are among the greatest challenges the world faces today. According to the World Bank<sup>1</sup>, air pollution has emerged as the fourth-leading risk factor for deaths worldwide.

Mobile operators can provide powerful and unique insights based on anonymised, aggregated network

data to help solve these complex problems. Mobility data can help public health organisations to more effectively respond to epidemics or plan targeted health interventions. It can support emergency relief agencies to more accurately and efficiently direct their resources. It can help governments better understand the impact of pollution and climate change on citizens.

Through the GSMA, mobile operators across geographies have come together to accelerate and scale the opportunity for Big Data for Social Good. The GSMA offers a unique platform to establish a common framework and best practice approaches, while respecting and protecting individuals' privacy. This project case study has been developed as part of the GSMA's Big Data for Social Good initiative, to promote similar models being rolled out to other major cities worldwide.

1. <http://www.worldbank.org/en/news/infographic/2016/09/08/death-in-the-air-air-pollution-costs-money-and-lives>

## Summary

Telefónica Brazil (Vivo) is working with the municipalities of São Paulo to harness mobile network data to help combat the adverse health impact of air pollution. Telefónica is providing anonymised mobile data, algorithms and tools, which the municipalities of São Paulo can integrate into their traffic and pollution management processes. Developed by the LUCA team, Telefónica's algorithms use machine learning and anonymised data from the mobile network, combined with data from weather, traffic and pollution sensors, to monitor and predict pollution levels over the entire city.

The solution can predict pollution levels 24 to 48 hours in advance, enabling local authorities in São Paulo to take preventative steps if nitrogen dioxide (NO<sub>2</sub>)

emissions could endanger human health. By employing mobile big data, Telefónica and the municipalities can significantly reduce costs, increase accuracy and provide valuable information for decision makers about how to define urban transport policies or assess pollution levels.

The implementation in São Paulo has demonstrated that mobile big data is a powerful tool with which to predict and analyse traffic patterns, and the resulting air pollution. Anonymised, aggregated mobile data allows for a more granular and cost-effective monitoring of pollution in the city compared with a static sensor network collecting equivalent data.

# The air pollution challenge

Multiple studies have shown the relationship between environmental pollution and public health issues. The World Health Organization has found air pollution is associated with chronic diseases and exacerbates related pathologies, such as cardiovascular and respiratory morbidity and mortality<sup>2</sup>. One study in Brazil<sup>3</sup> estimated more than 20,000 deaths a year, across 29 Brazilian metropolitan regions, can be attributed to air pollution.

As more environmental data becomes available, there is a global trend towards more data-driven approaches to managing air pollution and its impact, in line with the SDGs established by the United Nations:

- **SDG 3** Ensure healthy lives and promote well-being at all ages: SDG 3 seeks to increase life expectancy, partly through measures to strengthen the capacity of countries, in particular developing countries, for early warning, risk reduction and management of national and global health risks.
- **SDG 13** Take urgent action to combat climate change and its impacts.
- **SDG 11** Make cities and human settlements inclusive, safe, resilient and sustainable: SDG 11 seeks to reduce the adverse per capita environmental impact of cities, by 2030, partly by paying special attention to air quality and waste management.

Precision mapping of the levels of pollution across metropolitan areas can be used to:

- Identify potential local pollution peak levels that might affect the health of the local population, and, in the longer term, link these insights to other population data (e.g. asthma) to proactively inform citizens about the specific impact.
- Advise traffic management systems to reduce the number of vehicles in areas with high or dangerous pollution levels.

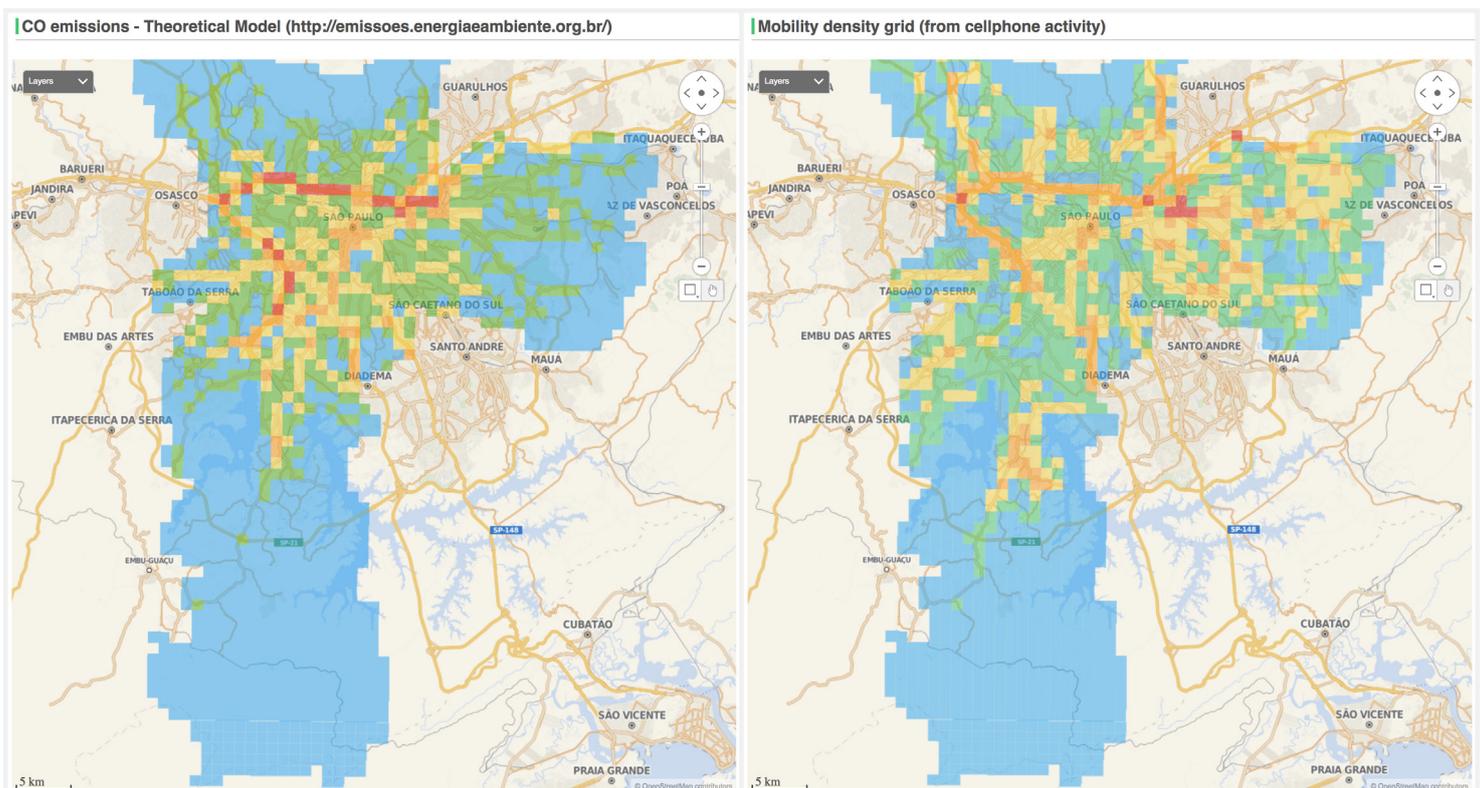
However, air quality management can be costly and traffic analysis can be relatively manual, using roadside interview data and manual counters. There is limited air quality monitoring infrastructure in most cities due to the high installation and operational cost of fixed measurement stations. This means data is sparse, making it difficult for public administrations to collect the accurate data they need to identify where the major air pollution hotspots are.

For instance, in São Paulo, the local government and municipalities' pollution maps are relatively imprecise, as they rely on a limited set of measurements. See example in Figure 1.

2. <http://www.who.int/airpollution/en/>

3. Miraglia S., Gouveia N. Costs of air pollution in Brazilian metropolitan regions. *Cienc. Saude Colet.* 2014;19:4141-4147

**Figure 1 Telefónica's dashboard shows the strong agreement between traffic measurements made from mobile data to the government's theoretical CO<sub>2</sub> predictions.**



## How mobile data can help

Telefónica Brazil is working with the municipalities of São Paulo to correlate information from air quality sensors, traffic sensors and weather stations with data generated by mobile networks. Such an approach can yield insights more cost-effectively than direct observation of pollution levels, and on a regular, more dynamic basis than is possible with more traditional data collection methods.

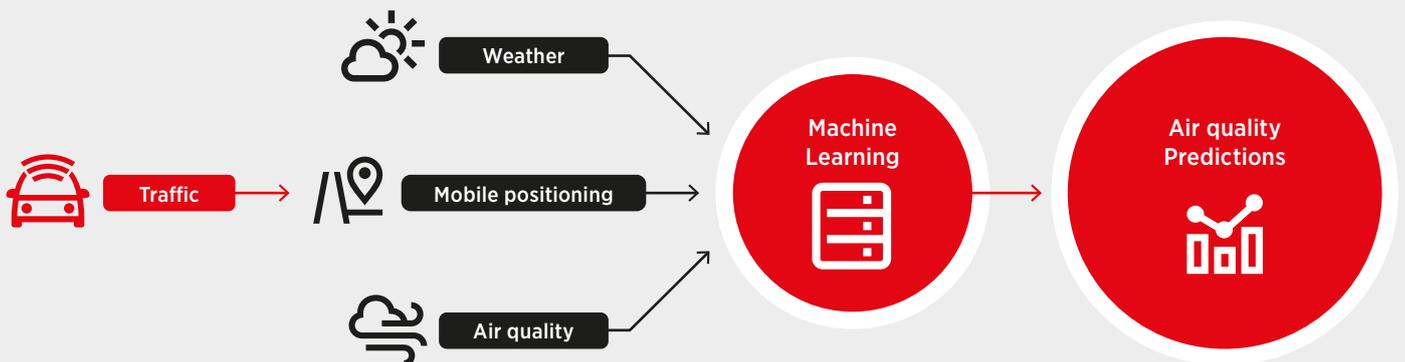
To compensate for the limited number of sensors across the city, Telefónica is leveraging mobility data to act as a proxy for observations about the volume and flows of vehicles. With its LUCA platform, Telefónica has created a set of algorithms to calculate mobility and traffic estimates, and a dashboard that displays pollution levels across the city. The algorithms use machine learning and weather, traffic and pollution data to extrapolate the mapping over the entire city (see Figure 2).

Telefónica's mobile data can encompass both active events, such as a voice call or SMS, and passive events, such as devices synchronising with cell towers. This data can be aggregated and scaled up into an index to represent the entire population of the city. According to Telefónica, such an index generally has been highly correlated (up to 94%) with the actual observed traffic. To safeguard individuals' privacy, anonymisation is achieved by removing the MSISDN, adding noise, aggregating, extrapolating from Telefónica's market share to the entire population and normalising the data<sup>4</sup>.

Telefónica is also testing the data solution to predict pollution in Madrid and create a robust approach that could be scaled to different geographies.

4. E.g. providing deviations from the mean mobility in each location - Telefónica compute the mobility index for each of the measurement stations. Correlation ranges from 89% to 97% depending on the station. The 94% is the average between the different correlations.

**Figure 2 Combining pollution data, traffic data and mobile data to map and predict air quality**



Mobile positioning data is used as a proxy for traffic data, allowing traffic to be measured across the whole city instead of at the locations of the limited network of traffic sensors. The mobile data, along with weather and historic air quality data, are combined in a machine learning model to generate air quality predictions.

## Impact of the mobile data solution

The implementation in São Paulo has proven mobile big data is a powerful tool with which to predict and analyse traffic patterns, and resulting air pollution, into and across a city. Overall, mobile data enables a more granular monitoring of pollution across the city than could be realised through a network of costly static sensors.

The São Paulo implementation highlights how mobile network data, correlated with air quality, weather and other data sources, can help to predict pollution

levels 24 to 48 hours in advance, enabling local authorities to take preventative steps if NO<sub>2</sub> emissions could endanger human health. Such an approach can significantly reduce costs, increase accuracy and provide valuable information for decision makers on how to define urban transport policies or assess pollution levels. Moreover, third parties could use an API to gain direct access to aggregated and anonymised telecoms data and mobility insights to meet their own needs.



## Developing a sustainable proposition

A key lesson learned is, to make this kind of implementation sustainable, local governments and mobile operators need to work together. By combining their datasets, they can create and analyse new models to generate better predictions and define potential policies based on these insights.

To fully unlock the potential of mobile big data, a key step is for public administrations to amend

their annual project budget to account for the management of these services and to invest in relevant skills. This will directly help to reduce the associated healthcare costs, local infrastructure development, and urban mobility. Some administrations are already setting aside some budget for data analytics and case studies, even specifically for mobile datasets to enhance their models<sup>5</sup>.

5. Innovation Plan for Transport & Infrastructure 2017 - Ministry of Public Works and Transport - Government of Spain

## Going forward

Telefónica has identified a series of potential next steps:



Enhance the tools in terms of robustness and relevance, e.g. leverage additional data sources, integrate them within the operational processes of the local authorities, and improve the quality of the predictive models.



Help municipalities to continue optimising the network of stations monitoring the levels of pollution and traffic across the cities and more broadly across the country to improve the algorithms and prediction capabilities.



Develop a roadmap to transform the São Paulo solution into more scalable products that could then be implemented in other geographies. Telefónica is developing a similar approach to cover the municipality of Madrid, the capital of Spain.



Further develop revenue streams from third parties to access mobile data aggregates through an API for other purposes requiring or involving mobility and/or environmental insights to create both economic (e.g. targeted marketing campaigns) and social (e.g. disaster preparedness) value.

Watch our video, learn about the initiative and contact us for more information: [bd4sg@gsma.com](mailto:bd4sg@gsma.com)

[gsma.com/betterfuture/bd4sg](https://gsma.com/betterfuture/bd4sg)



**@GSMA**

**#BetterFuture #BD4SG**