External exposome enrichment

Methodological aspects (LISA)

Work from the **LISA team** during the REMEDIA project focused on developing and providing several relevant statistical indicators of air quality for the **VLM** and **PELAGIE** cohorts, which have different patient location data and do not require the same resolution of air quality data. **The general approach** of LISA in the REMEDIA project was to enrich air quality data in patient cohorts, firstly by adding a better spatial representation of individual exposure risk, and by adding to conventional exposure data (annual concentration averages) innovative indicators capturing several specific dimensions of each individual's exposure. We relied on robust data from the Chimere chemistry-transport model at a fine spatial resolution. First, we developed an exposure attribution code to discriminate between proximity to traffic in the living environment of individuals, which is a determining factor in the burden of urban exposure. We applied a statistical modulation of the signal based on learning approaches to reproduce the increment in exposure that stems from traffic proximity. Then, designed a set of exposure indicators reproducing patterns such as the frequency and intensity of exposure above WHO threshold values, for a set of regulatory pollutants (NO_2, O_3, $PM_{2.5}$ and PM_{10}).

Exposure calculation

To best characterize the exposure of individuals in the cohorts (when the home address is available), it is important to determine the typology of the living environment, distinguishing between urban, suburban, rural, etc. environments. To do this, we have developed a classification method with a spatial resolution of 1km, based on the criteria used to determine the nature of a measuring station (urban background, urban traffic, rural background, etc.) at national level (LCSQA). This method is based on the cross-referencing of several parameters: number of inhabitants, typology of urban areas and land use. This enables us to propose a classification into 5 categories: urban, peri-urban, rural, airports and ports, which will then be used to calculate exposure.

Depending on proximity to sources, exposure to pollutants can vary greatly. In the city, for example, it will be higher in a street with a lot of traffic than in a park. However, the CTM is unable to reproduce the urban gradients observed, particularly near road traffic. In response to this shortcoming in the model, we have worked on a statistical modulation of the signal (based on observation data) that corrects the modelled data to better reproduce urban road environments.

- **Application to the VLM cohort -** The individual data provided for the VLM cohort are at departmental level. To meet this constraint, we therefore had to aggregate the concentration fields provided by CHIMERE (which have a resolution of between 3 and 5 km) at the scale of the departments. We thus obtained daily average concentrations for each department over the periods covered by the modelling data. Modelling data are available for the 4 regulated pollutants: NO_2,

O_3, $PM_{2.5}$ and PM_{10}, and cover different time periods. For NO_2 and O_3, the data is available between 2000-2019, for PM_{10}, the data covers the period 2007-2019 and for $PM_{2.5}$ the data is available over the period 2009-2019.

- **Application to the PELAGIE cohort -** For the PELAGIE cohort, we have the location of individuals' homes throughout their follow-up period (so we can have several addresses), but we also have information on the dates of follow-up with physicians. Concentrations for this cohort can therefore be assigned in a more detailed way, allowing us to apply concentrations taking into account the type of living environment (urban, rural) and proximity to roads. For each individual in the cohort we therefore determined the type of home environment, then depending on its location and proximity to a major road, we assigned a daily concentration according to several criteria as described in the figure below. We then calculated the indicators for each life period of the individuals in the cohort (pregnancy, childhood, etc.).

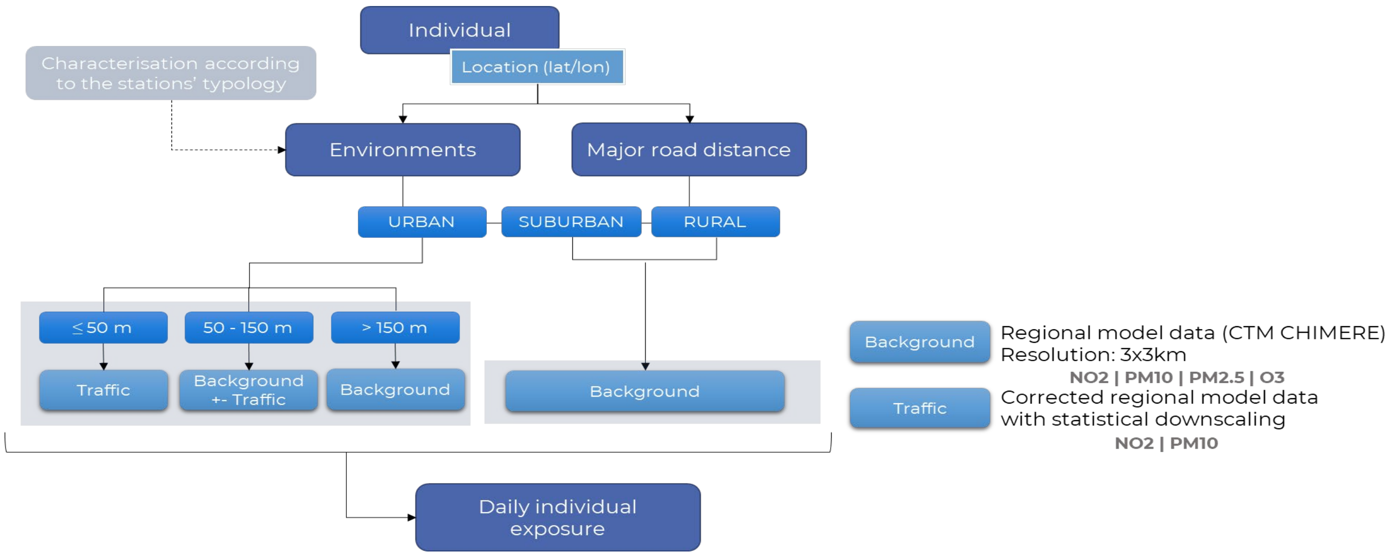

Figure 2.4: *Patterns of exposure through statistical indicators*

The first step was to propose exposure indicators that could best discriminate between individuals, to assess the impact of exposure on health trajectories and to reproduce heavy or chronic exposure. We therefore selected the 90 percentile (P90) of daily values (in $\mu g/m^3$), the number of days over the year when the WHO2021 threshold was exceeded, and the cumulative difference between the daily value and the WHO threshold (when positive, in $\mu g/m^3$).

These indicators can be calculated at different spatial scales, depending on the resolution of the air quality data available. Thus, for the VLM cohort, the indicators were calculated at departmental level, while for the PELAGIE data they were calculated at the level of the individual's home address.