

Spatial interpolation techniques for innovative air quality monitoring systems

Nicoletta Lotrecchiano^{a,b}, Daniele Sofia^{a,b*}, Aristide Giuliano^c, Diego Barletta^a, Massimo Poletto^a

^a DIIN-Dipartimento di Ingegneria Industriale, Università degli Studi di Salerno, Via Giovanni Paolo II 132, 84084, Fisciano (SA), Italy

^b Sense Square srl, Corso Garibaldi 33, 84123, Salerno (SA), 84123, Italy

^c ENEA, Italian National Agency for New Technologies, Energy and Sustainable Economic Development, Laboratory Biorefineries and Green Chemistry, S.S. 106 Ionica, km 419+500, Rotondella, MT, Italy

*dsofia@unisa.it

New technologies for air quality measurement, including real-time on-road air quality monitoring systems (ROMs), have improved the spatial resolution of data. However, measured data can miss due to restricted traffic zones. Different algorithms can estimate the missing data in these points, using available measured data in the neighbourhood. In this study, the results of the applications of an interpolation method taking into account the effect of wind direction and intensity by means of a simple dispersion model is presented. Thanks to the large amount of data provided by the innovative dynamic monitoring system (ROM), the proposed model is able to return, with a good approximation, reliable PM10 and PM2.5 concentration values, with a resolution of 1 km².

1. Introduction

In recent years air quality data are used to investigate the relationship between air pollution and human health to quantify the effect of exposure to pollutants (Donaire-Gonzalez et al., 2019). With this regard, the European Union reported daily and annual limit concentrations for some of the major pollutants in the Directive 2008/50/EC to support decarbonisation (Sofia et al., 2020a). Therefore, it is necessary to create accurate monitoring networks, which can measure and record time series of major pollutant concentrations (Sofia et al., 2018). Analysis of these data allows the pollution forecast (Lotrecchiano et al., 2019) and investigating the effect of pollution sources by means of dispersion models at steady state (Sofia et al., 2020b) and over the time depending on the weather conditions and emission factors (Lotrecchiano et al., 2020). The accuracy of models depends on many factors, including the spatial and time resolution of the source data, the quality of the meteorological data in the area, the assumptions about the physical and chemical processes in the atmosphere involving the transport and conversion of pollutants (Sofia et al., 2020c).

Recently, cheaper smart measurement devices (Sofia et al., 2019) allow the measurement of pollutant concentrations with higher spatial-temporal resolution. Moreover, the design and implementation of portable measuring devices, like the real-time on-road monitoring stations (ROMs) (Lotrecchiano et al., 2019), allows collecting a huge amount of data with a much higher spatial resolution. However, a lack of data may occur in areas inaccessible to vehicles, such as traffic-restricted zones or city parks. Spatial interpolation methods based on either a deterministic or stochastic approach can be used to estimate missing data. In addition to simple linear interpolation, kriging is one of the most used statistical methods for physical data depending on atmospheric dispersion in Geostatistics, the science studying the natural phenomena that develop on the territory. In particular, kriging is a spatial interpolation method combining an atmospheric dispersion model and a pollutant emissions inventory to interpolate the model outputs (Beauchamp et al., 2018). Differently, Inverse Distance Weighting (IDW) calculates missing concentration data based on the weighted sum of the neighbouring observations considering the inverse of the distance. In general, deterministic methods do not take into account the variability in the spatial domain of the parameter to be interpolated and their implementation is arbitrary.

Furthermore, the results are extremely dependent on the spatial configuration of the input data and on the sampling schemes. The measures provided by the monitoring systems in general are sparse and must be expanded in space through the implementation of a model to have clear information on the environmental situation.

In this work air quality data measured by ROMs over a very large area, the city of Milan, are elaborated to find a suitable interpolation procedure to calculate pollutant concentrations in positions where measurement data are not available. It is expected that the availability of a large amount of data distributed on the ground could allow to obtain significant interpolation results using fairly simple approach to pollution data.

2. Materials and methods

Data

The air quality data used in the present work were measured by the ROMs network installed in Milan (Italy). This type of technology is patented and properly designed to be located on moving vehicle. The network consists of 53 measuring devices located on courier vans. Each device provides measurements of the main aero-dispersed pollutant concentration such as particulate matter of different sizes PM1, PM2.5, PM10, and gases as H₂S, CO, O₃, NO₂, SO₂ and, VOC. In addition to the air quality measurements, meteorological data as temperature, pressure and relative humidity are acquired as well. Wind intensity and direction data are obtained from an external meteorological station using an application programming interface key (API key). The ROMs data are protected and unchangeable due to the implementation of a blockchain system that guarantee the data integrity (Sofia et al., 2020d). The area covered by the moving measuring stations is about 306 km² and corresponds to the whole metropolitan area of the city of Milan. In this work only the PM10 and PM2.5 concentration data are used. The measuring device provides latitude and longitude coordinates via GPS. The data acquired every 3 minutes are aggregated on a daily base and with a spatial resolution of 1 km². Data are reported on a grid of 1 km² square cell and the data acquired in January and February 2020 were used in this work.

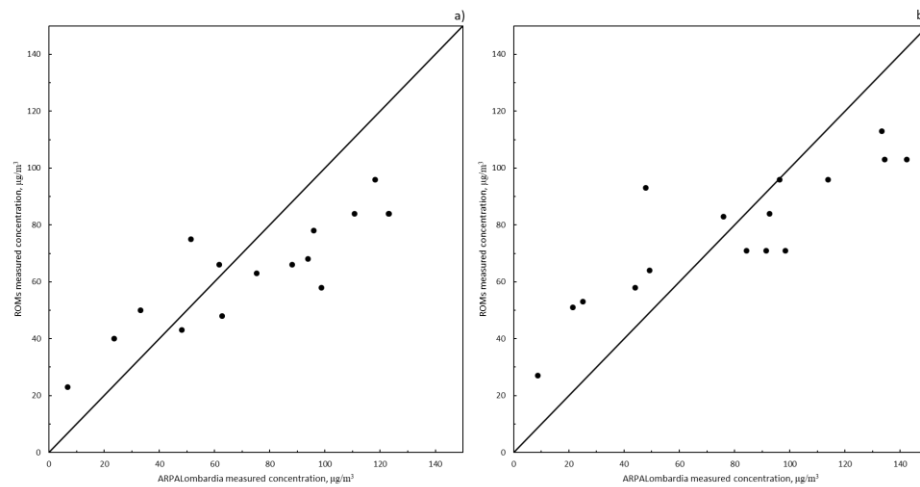


Figure 1. Comparison between PM10 daily concentrations measured by the ROMs and the ARPALombardia air quality network in a) Verziere and b) Senato in Milan on January 2020.

The ROMs data was compared with the measured one provided by the ARPA Lombardia air quality network, which provides pollution data collected according to the Italian Legislative Decree 155/2010. In Figure 1 it is clear that the two measuring technologies (laser scattering and gravimetry) provides data fairly well comparable.

Methodology

Raw concentration data, c , are transformed into deviation with respect to an assumed background value, c_{bg} , corresponding to the minimum concentration value recorded in the whole grid. Therefore, the interpolation procedure processes the deviation values $C=(c-c_{bg})$. In order to estimate missing concentration data, each cell, where measured data is available, can be considered that pollution data in neighboring cells are interdependent and that the interdependence decreases with the distance but is stronger in the wind direction as it may happen

if each point is considered as a pollutant emission source. This could be more detailed, if it could be possible to identify the specific intensity of emission source in each area. However, in an urban area with roads, houses and other activities as a source of particulate emissions an assumption of a fairly uniform emission on the ground is a good approximation of the real situation. In the following the procedure adopted to evaluate the mutual cell interdependence. Each measured data in a cell is considered the starting point of the spatial expansion of locally produced pollutants and provides its contribution to an expansion matrix. After the application of the expansion model to all cell with a measured value, the obtained matrices are combined to obtain a final matrix. The latter is given by linear combination of all the estimated matrices with a weight proportional to the distance from the emission source cell. The spatial expansion takes into account the wind intensity and direction to model the pollutant dispersion phenomenology. The approach used in this study assumes a pollutant dispersion field such that each emission source could generate iso-concentration elliptical lines, with the source cell centre located in one of the focal points. The concentration in the points of each ellipse is equal to the measured concentration in the source cell multiplied by a decay factor depending on the ellipse size. It is assumed that the main axis of each ellipse is placed along the wind direction and the ratio between the ellipse axes is a function of the wind speed. Accordingly, the axes have been defined by the following equations (Eq.1):

$$\begin{cases} a^2 + b^2 = 2r^2 \\ a^2 = b^2 + \gamma r^2 v^2 \end{cases} \quad (1)$$

where a is the semi-major axis, b is the semi-minor axis, v is the normalized wind speed, r is the distance at which a given concentration decay coefficient is obtained in the absence of wind.

The estimated concentration in each cell in (i, j) position, $CE_{i,j}$, is a linear combination of the concentrations of the neighbouring source cells, $C_{k,l}$, properly weighted by means of factors corresponding to the source cells ellipses, passing through the (i, j) cell itself. In particular, the considered source cells are the closest eight on the grid placed around the (k, l) cell. The resulting concentration is given by the Eqs. (3) and (4):

$$CE_{i,j} = \sum_{\substack{k=i-1,i+1 \\ k \neq i}} \sum_{\substack{l=j-1,j+1 \\ l \neq j}} C_{k,l} w_{k,l} \quad (2)$$

$$w_{k,l} = \frac{\alpha_{k,l}}{\sum_{\substack{k=i-1,i+1 \\ k \neq i}} \sum_{\substack{l=j-1,j+1 \\ l \neq j}} \alpha_{k,l}} \Rightarrow \sum_{\substack{k=i-1,i+1 \\ k \neq i}} \sum_{\substack{l=j-1,j+1 \\ l \neq j}} w_{k,l} = 1 \quad (3)$$

where $w_{k,l}$ is the weighing factor of the concentration contribution for the neighbouring cell in the (k, l) position and $\alpha_{k,l}$ is the corresponding decay coefficient. The decay coefficient was evaluated in two alternative ways. In the first case, $\alpha_{k,l}$ decreases linearly with the size $r_{k,l}$ of the ellipse, having a focus in the point in the (k, l) position and passing through the point in the (i, j) position, and defined by Eq (4):

$$\alpha_{k,l} = 1 - r_{k,l} d_c \quad (4)$$

where d_c is a decay constant with $d_c=0.05$. In the second case, $\alpha_{k,l}$ is evaluated by an exponential decay function with $d_c=1.57$:

$$\alpha_{k,l} = \exp(-r_{k,l} d_c) \quad (5)$$

The parameter γ appearing in Eq. (1) is obtained by minimizing the deviation function, f , obtained as the sum of the squares of the residuals between the estimated concentration $CE_{i,j}$ and the measured concentration $C_{i,j}$, for the neighbouring cells where measured data are available:

$$f = \sum_k \sum_l \sqrt{(CE_{i,j} - C_{k,l})^2} \quad (6)$$

To evaluate the performance of the complete model, the normalized root mean square error between the measured values and the corresponding values estimated with the model was calculated as follows:

$$NRMSE = \sqrt{\frac{1}{N} \sum_k \sum_l \left(\frac{CE_{i,j} - C_{k,l}}{C_{k,l}} \right)^2} \quad (7)$$

where N is the number of the values estimated.

3. Results

Preliminary calculations aimed at evaluating the optimal parameters and equations of the spatial dispersion model. Figure 2b shows that the f function, defined by Eq. (6), is weakly dependent on the parameter γ , probably due to the normalization of Eq. (3). Nevertheless, a minimum value can be observed and, thus, an optimal value

for γ was set to 1.5. The parity plot of Figure 2a compares the concentration values, estimated by Eq. (2), using for the decay coefficient $\alpha_{k,l}$ Eq. (4) or Eq. (5), respectively. Inspection of the figure reveals that there is not a significant difference between the estimated values. As a result, the exponential decay function of Eq. (5) was adopted for the rest of the study since it yields intrinsically $\alpha_{k,l}$ values between 0 and 1.

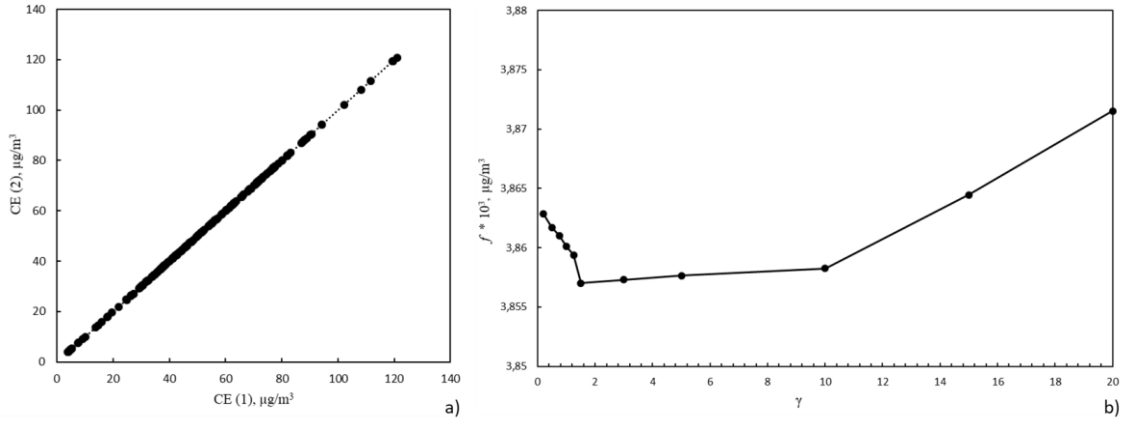


Figure 2. 2a) Comparison between CE concentration values estimated by Eq.(2): CE (1) with $\alpha_{k,l}$ according to Eq.(4), CE (2) with $\alpha_{k,l}$ according to Eq.(5). 2b) Minimization of f as a function of parameter γ .

The grid (x, y) covering the investigated area of Figure 1, also reported in Figure 3, was rotated by an angle β corresponding to the wind direction. Figure 4 reports the eight ellipses with constant decay coefficient values, calculated according to Eq. (5), passing through a selected (i, j) point, to estimate $CE_{i,j}$ as a function of the measured concentration in the eight neighbouring cells. Inspection of the four panes, corresponding to different normalized wind intensity values, suggests that the higher is v , the longer is the ellipse major axis and the more effective is the pollutant dispersion. This scenario corresponds to lower $\alpha_{k,l}$ values for the source cells located downwind with respect to the (i, j) point and to higher $\alpha_{k,l}$ values for the source cells located upwind. As a result, the concentration $CE_{i,j}$ is affected by the concentration of the neighbouring cells with different weights. On the contrary, the lower is v , the shorter is the ellipse major axis and the less effective is the pollutant dispersion. In absence of wind (Figure 5a), i.e. $v = 0$, the ellipses turn into circles corresponding to the lowest dispersion and the pollutant concentration tends to accumulate in the vicinity of the source cells. In this case, comparable $\alpha_{k,l}$ values are obtained for all the neighbouring cells and their corresponding concentration values affect the concentration $CE_{i,j}$ with the same weight.

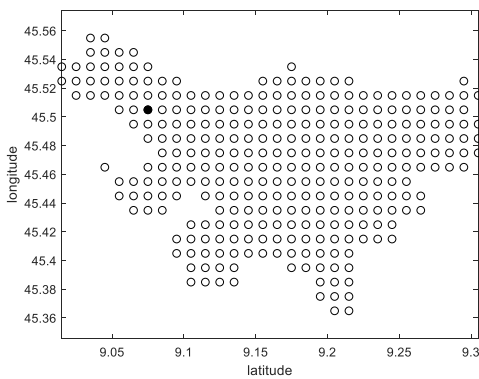


Figure 3. Example of point location on the measurement position map.

Figure 5 shows an example of the resulting PM10 and PM2.5 maps after the spatial interpolation according to Eq. (3) and using the measured concentration data, the wind intensity and direction. Concentration is represented according to a colour scale. Solid circles correspond to measured data, while hollow circles correspond to estimated data. The estimated values appear in good agreement with the close measured data.

The estimate is less satisfactory in the periphery of the map where less measured points were available and the background concentration value c_{bg} was assumed. On the whole, the interpolation model appears effective considering that the error function $NRMSE$ values, obtained according to Eq. (7), were in the range of 0.08-0.7 for the majority of tested concentration maps. Analyzing the $NRMSE$ calculated using the estimated data closest to the regional monitoring devices, the value of 0.1 is obtained, indicating a good agreement between the estimate and the measured values. Comparison with other spatial estimation methods will be the subject of further work.

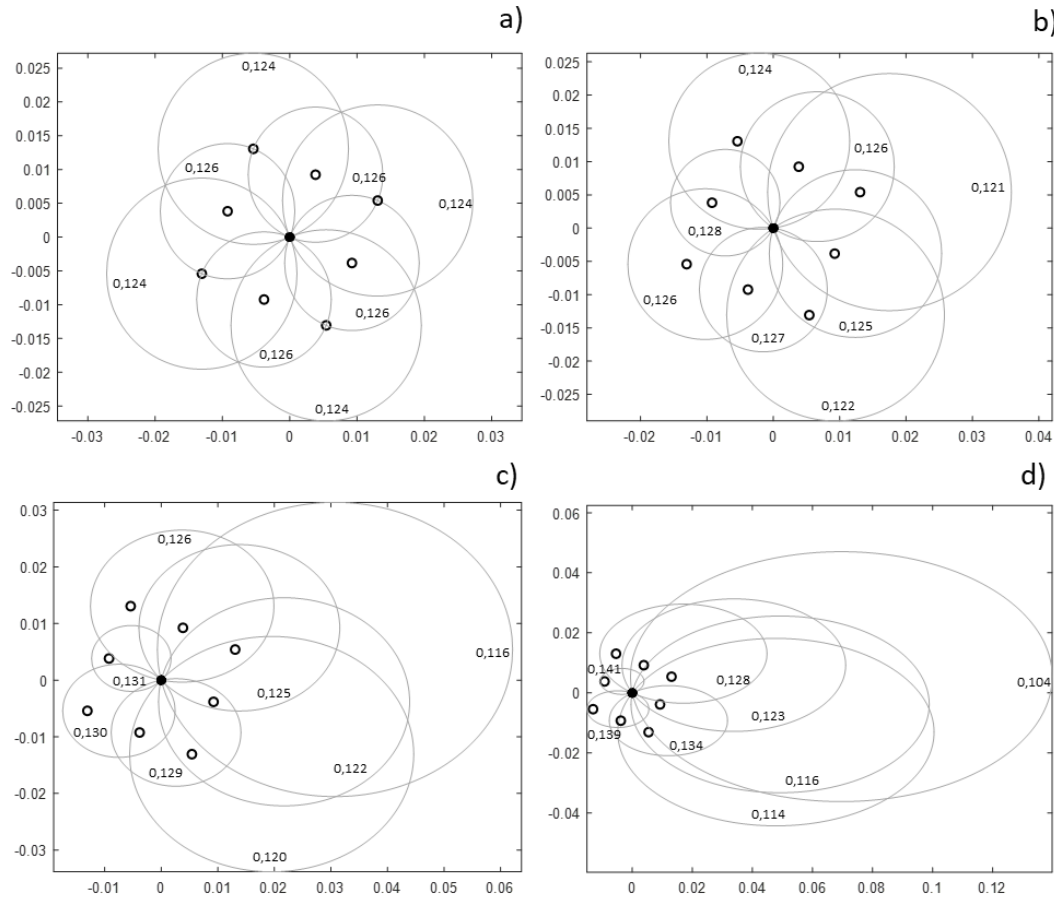


Figure 4. Decay coefficient elliptical fields to evaluate the concentration in the (i, j) point corresponding to the solid circle as a function of the concentrations measured in the eight neighbouring (k, l) points represented by the open circles for a) $v=0$, b) $v=0.2$, c) $v=0.5$, d) $v=0.8$ at wind direction of 247.5° . Corresponding $\alpha_{k,l}$ values are reported on the curves.

4. Conclusions

The proposed interpolation method appears to describe qualitatively the effect of wind direction and intensity on the pollutant dispersion. The parameter relating the wind speed and the decay coefficient function $\alpha_{k,l}$ was optimized. This parameter plays a more important role than the mathematical function type chosen for $\alpha_{k,l}$ on the interpolation performance, provided that the weighing factors $w_{k,l}$ of the neighbouring concentrations are normalized. The resulting deviation between measured values of PM10 and those estimated by the proposed method is generally lower than $4 \mu\text{g}/\text{m}^3$ and only in rare cases it approaches $10 \mu\text{g}/\text{m}^3$ with $NRMSE$ of about 0.3. These satisfactory results provide some confidence in the modelling approach and encourage further developments. Future work will consider a possible increase of the number of points with measured data to estimate the concentration in peripheral points. Finally, this type of spatial expansion could be used also in combination with other dispersion approaches like the Gaussian Plume and Puff dispersion models.

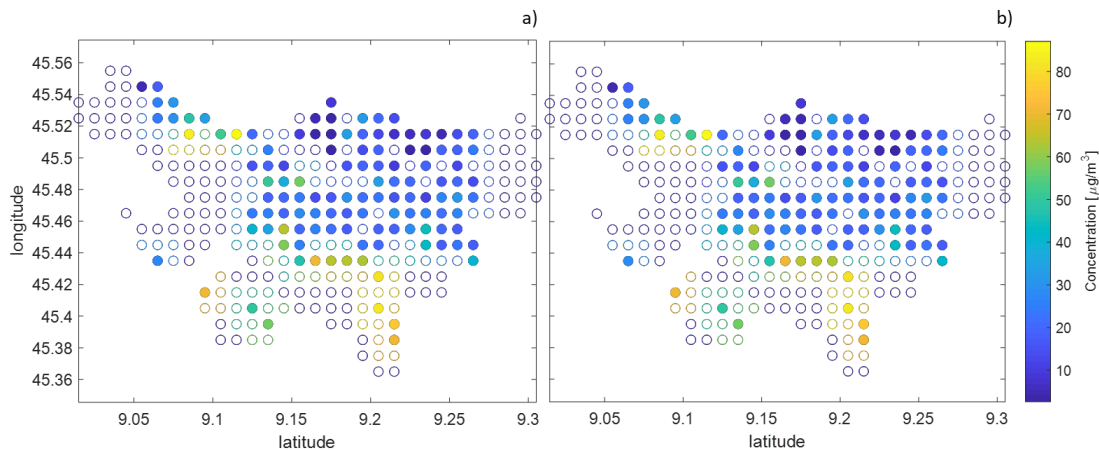


Figure 5. a) PM10 and b) PM2.5 concentrations estimated applying the spatial interpolation model. Solid circles correspond to measured data while hollow circles correspond to estimated data.

References

- Beauchamp M., Malherbe L., de Fouquet C., Létinois L., Tognet F., 2018, A polynomial approximation of the traffic contributions for kriging-based interpolation of urban air quality model, *Environmental Modeling Software*, 105, 132–152, <https://doi.org/10.1016/j.envsoft.2018.03.033>
- Donaire-Gonzalez D., Valentín A., van Nunen E., Curto A., Rodriguez A., Fernandez-Nieto M., Naccarati A., Tarallo S., Tsai M.-Y., Probst-Hensch N., Gulliver J., Nieuwenhuijsen M.J., 2019, ExpoApp: An integrated system to assess multiple personal environmental exposures, *Environment International*, 126, 494–503, <https://doi.org/10.1016/j.envint.2019.02.054>
- Lotrecchiano N., Sofia D., Giuliano A., Barletta D., Poletto M., 2020, Pollution dispersion from a fire using a Gaussian plume model, *International Journal of Safety Security Engineering*, 10, 431–439, <https://doi.org/10.18280/ijssse.100401>
- Lotrecchiano N., Sofia D., Giuliano A., Barletta D., Poletto M., 2019, Real-time on-road monitoring network of air quality, *Chemical Engineering Transactions*, 74, 241–246, <https://doi.org/10.3303/CET1974041>
- Lotrecchiano N., Gioiella F., Giuliano A., Sofia D., 2019, Forecasting Model Validation of Particulate Air Pollution by Low Cost Sensors Data, *Journal of Model Optimization*, 11, 63–68, <https://doi.org/https://doi.org/10.32732/jmo.2019.11.2.63>
- Sofia D., Gioiella F., Lotrecchiano N., Giuliano A., 2020a, Mitigation strategies for reducing air pollution, *Environmental Science Pollution Research*, 27, 19226–19235, <https://doi.org/10.1007/s11356-020-08647-x>
- Sofia D., Gioiella F., Lotrecchiano N., Giuliano A., 2020a, Cost-benefit analysis to support decarbonization scenario for 2030: A case study in Italy, *Energy Policy*, 137, 111137, <https://doi.org/https://doi.org/10.1016/j.enpol.2019.111137>
- Sofia D., Giuliano A., Gioiella F., 2018, Air quality monitoring network for tracking pollutants: The case study of salerno city center, *Chemical Engineering Transactions*, 68, 67–72, <https://doi.org/10.3303/CET1868012>
- Sofia D., Lotrecchiano N., Cirillo D., Villetta M.L., Sofia D., 2020b, NO2 Dispersion model of emissions of a 20 kwe biomass gasifier, *Chemical Engineering Transactions*, 82, 451–456, <https://doi.org/10.3303/CET2082076>
- Sofia D., Lotrecchiano N., Giuliano A., Barletta D., Poletto M., 2019, Optimization of number and location of sampling points of an air quality monitoring network in an urban contest, *Chemical Engineering Transactions*, 74, 277–282, <https://doi.org/10.3303/CET1974047>
- Sofia D., Lotrecchiano N., Trucillo P., Giuliano A., Terrone L., 2020d, Novel Air Pollution Measurement System Based on Ethereum Blockchain, *Journal of Sensors and Actuator Networks*, 9, 49, <https://doi.org/10.3390/jsan9040049>