Environment International 121 (2018) 57–70

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Bayesian geostatistical modelling of PM_{10} and $PM_{2.5}$ surface level concentrations in Europe using high-resolution satellite-derived products

Anton Beloconi[a,b], Nektarios Chrysoulakis[c], Alexei Lyapustin[d], Jürg Utzinger[a,b], Penelope Vounatsou[a,b,*]

[a] Swiss Tropical and Public Health Institute, Basel, Switzerland
[b] University of Basel, Switzerland
[c] Foundation for Research and Technology - Hellas, Heraklion, Greece
[d] NASA Goddard Space Flight Center, Greenbelt, MD, USA

ARTICLE INFO

Handling Editor: Xavier Querol

Keywords:
Particulate matter
Bayesian geostatistics
Integrated nested Laplace approximation
Aerosol optical depth
MAIAC
Copernicus

ABSTRACT

Air quality monitoring across Europe is mainly based on *in situ* ground stations, which are too sparse to accurately assess the exposure effects of air pollution for the entire continent. The demand for precise predictive models that estimate gridded geophysical parameters of ambient air at high spatial resolution has rapidly grown. Here, we investigate the potential of satellite-derived products to improve particulate matter (*PM*) estimates. Bayesian geostatistical models addressing confounding between the spatial distribution of pollutants and remotely sensed predictors were developed to estimate yearly averages of both, fine ($PM_{2.5}$) and coarse (PM_{10}) surface *PM* concentrations, at 1 km^2 spatial resolution over 46 European countries. Model outcomes were compared to geostatistical, geographically weighted and land-use regression formulations. Rigorous model selection identified the Earth observation data which contribute most to pollutants' estimation. Geostatistical models outperformed the predictive ability of the frequently employed land-use regression. The resulting estimates of PM_{10} and $PM_{2.5}$, which represent the main air quality indicators for the urban Sustainable Development Goal, indicate that in 2016, 66.2% of the European population was breathing air above the WHO air quality guidelines thresholds. Our estimates are readily available to policy makers and scientists assessing the effects of long-term exposure to pollution on human and ecosystem health.

1. Introduction

The contribution of particulate matter (*PM*) concentration to air pollution and the effects of high levels of these pollutants to human health and wellbeing have been documented extensively in the literature. Exposure to high concentrations of *PM* has been associated with increased rates of morbidity and mortality, caused primarily by cardiovascular, respiratory and, to a lesser extent, cerebrovascular diseases (Anderson et al., 2012).

Although a relatively dense air quality monitoring network exists in Europe, maintained by the European Environment Agency's (EEA) member states, large areas within the continent remain unmonitored. A number of approaches have been used to provide gridded pollutants' concentration estimates. On European and global scale, these include empirical models based on chemistry transport model outputs (van Donkelaar et al., 2006), land-use regression (LUR) (Beelen et al., 2009; Vienneau et al., 2013), kriging (Beelen et al., 2009) and geographically weighted regression (GWR) (van Donkelaar et al., 2016).

Unlike the aforementioned implementations, Bayesian inference allows the uncertainty in predictions to be assessed and taken into account in further analyses. The assessment of exposure burden by utilizing high-resolution population estimates becomes straightforward, in a way that is not possible in traditional approaches, since full posterior predictive distributions can be derived. However, predictions of pollutant levels at high-spatial resolution over the entire Europe are computationally complex (Shaddick et al., 2013). The computational burden can be reduced through approximate Bayesian inference using integrated nested Laplace approximation (INLA) (Rue et al., 2009; Lindgren et al., 2011). The potentials of this approach have been demonstrated on PM_{10} data covering a small study area in northern Italy (Cameletti et al., 2013).

Most of the data-driven air-quality assessments incorporate

* Corresponding author at: Swiss Tropical and Public Health Institute, Basel, Switzerland.
E-mail addresses: anton.beloconi@swisstph.ch (A. Beloconi), zedd2@iacm.forth.gr (N. Chrysoulakis), Alexei.I.Lyapustin@nasa.gov (A. Lyapustin), juerg.utzinger@swisstph.ch (J. Utzinger), penelope.vounatsou@swisstph.ch (P. Vounatsou).

https://doi.org/10.1016/j.envint.2018.08.041
Received 28 February 2018; Received in revised form 16 August 2018; Accepted 17 August 2018
0160-4120/

geographical covariates derived from satellite-based observations. Remotely sensed products provide spatial coverage over the entire domain of interest allowing regular monitoring of the pollutants' spatial distributions. The main satellite-derived product used for the estimation of surface *PM* concentration is the aerosol optical depth (AOD), which represents the integrated radiation scattering and absorption by aerosols in an atmospheric column from the surface to the top of the atmosphere. AOD is used as a proxy for *PM* since it depends on the mass concentration and size distribution of the particles. A number of methods have been developed for near-surface *PM* estimation using columnar AOD (Chu et al., 2016). Large geographical scale predictions are usually based on Moderate Resolution Imaging Spectroradiometer (MODIS) Dark Target AOD (Levy et al., 2013) available at $\sim 10\,\text{km}^2$ spatial resolution. The recently developed multi-angle algorithms for AOD retrievals using spaceborne observations such as the Medium Resolution Imaging Spectrometer (MERIS)/Advanced Along-Track Scanning Radiometer (AATSR) (North et al., 2009) and the Multi-Angle Implementation of Atmospheric Correction (MAIAC) (Lyapustin et al., 2011) algorithms provide AOD distributions at $1\,\text{km}^2$ spatial resolution rendering the product suitable for high resolution *PM* modelling (Chudnovsky et al., 2014; Hu et al., 2014; Beloconi et al., 2016). However, there are many challenges in predicting *PM* conditioning on observed AOD. Due to the vertical structure of AOD, the strength of the *PM*-AOD relationship varies greatly in both space and time (Lee et al., 2011; Hu et al., 2014). Thus, it was shown that MODIS Dark Target AOD provides little additional information in a model that already accounts for local emissions, meteorology, land-use and regional variability at monthly and annual averaged temporal resolution level (Paciorek and Liu, 2009).

The primary objective of this work was to assess the benefits of combining satellite-derived products in a rigorous geostatistical modelling framework to estimate pollutants' spatial variability over 46 European countries. Particularly, the contribution of the MAIAC aerosol information adjusted with a set of georeferenced predictors, including the novel Copernicus land products (Copernicus, 2018) and meteorological data was evaluated for estimating high-resolution ($1\,\text{km}^2$) pollutant maps of PM_{10} and $PM_{2.5}$ using hierarchical Bayesian spatial models. We compared different model formulations and assessed the effect of confounding between spatially varying predictors and the spatial process, which incorporates geographical correlation in the pollutants' concentration. Furthermore, the Bayesian formulation allowed us to quantify the prediction uncertainty, to determine at high spatial resolution areas that exceed the European Union (EU) and World Health Organization (WHO) air quality guidelines' (AQG) thresholds as well as to estimate the number of people living in such areas. Model fit was done using the INLA algorithm. The models provide improved gridded air-quality estimates for policy makers and scientists assessing the effects of pollution on human and ecosystem health.

2. Materials and methods

2.1. Study area and data

The PM_{10} and $PM_{2.5}$ data were obtained from the Air Quality e-Reporting database (Air Quality e-Reporting, 2018) maintained through the Eionet (European environment information and observation network). The monitoring network covers up to 38 European countries, including the 28 EU member states and 33 member countries of the EEA. The repository consists of a multi-annual hourly time series data for a list of air pollutants. In this work the analysis is based on the yearly averaged data (reported in $\mu g/m^3$) of 2016 (currently the most recent year with available raw data) at stations with $\geq 75\%$ data capture. Fig. 1 (a–b) illustrates the locations of the monitoring sites used in this work, together with the yearly averaged measured concentrations of PM_{10} and $PM_{2.5}$. All data used in our analyses were converted to the Lambert Azimuthal Equal Area (ETRS89-LAEA5210) projection recommended by the EEA (European Environment Agency, 2006) for storing raster data, statistical analysis and map display purposes.

The satellite-derived product of columnar aerosol optical depth was considered as proxy of surface *PM* concentration. The recently-developed MAIAC algorithm for aerosol retrievals is based on time-series analysis and image processing of MODIS satellite data. MAIAC uses empirically tuned, regionally prescribed, aerosol properties following the AERONET (AErosol RObotic NETwork) climatology and provides AOD values over land at $1\,\text{km}^2$ spatial resolution globally. A total number of 19 tiles from each of the MODIS Terra and Aqua satellites, covering the study area (ftp://dataportal.nccs.nasa.gov/DataRelease/Europe_Sept-2017/), were downloaded and preprocessed. Each image was reprojected to our study area using the MODIS Reprojection Tool (MRT, 2016) accessed through the R environment (R Core Team, 2015). The raster package (Hijmans, 2015) was used to mosaic the resulting products. To reduce the number of the missing pixels, first daily, and then yearly averages from two satellites (i.e. Terra and Aqua) were computed.

Several studies have evaluated the effects of land-use/cover, urban mapping, local climate and meteorology information on the estimation of both PM_{10} and $PM_{2.5}$ (e.g. Liu et al., 2005, 2009; Benas et al., 2013; Vienneau et al., 2013; Chudnovsky et al., 2014; Stafoggia et al., 2016; He and Huang, 2018). These parameters influence the relationship between AOD and *PM* and can be used as predictors in assessing the geographical variation of pollutants' concentration. In order to be able to estimate *PM* over the whole Europe, we analysed satellite-derived products only with continental or global coverage. Table 1 summarizes the covariates used in this work. On its own, this data portfolio represents a powerful resource for numerous environmental applications in Europe.

The land-use/cover data were extracted from the pan-European component of the Copernicus Land Monitoring Service (Copernicus, 2018). For the temporal alignment with the observations from stations the latest CORINE Land Cover (CLC) dataset (year 2012) was used. To better understand the urban surface characteristics that surround the monitoring stations, a squared buffer zone of $1\,\text{km}^2$ spatial resolution around each station was created and the dominant CLC category within each buffer zone was computed and assigned to the respective site. The 45 land classes available in CLC were aggregated to form the following 4 main categories: (i) continuous urban fabric - road and rail networks and associated land - port areas (LC1); (ii) discontinuous urban fabric - industrial or commercial units - mine, dump and construction sites - artificial, non-agricultural vegetated areas (LC2); (iii) agricultural areas - wetlands - water bodies (LC3); and (iv) forest and semi-natural areas (LC4). Thus, it was expected that the pollution levels gradually decrease for stations situated in LC2–LC4 categories compared to the LC1, considered as the baseline.

Additionally, the high resolution layers of tree cover density (TCD) and imperviousness (IMP), as well as the European settlement map (ESM) were accessed from the same source (Copernicus, 2018). The PM_{10} and $PM_{2.5}$ levels were expected to be higher in build-up areas and lower in zones with higher tree density and therefore less emission sources. The digital elevation model (DEM) product (DEM, 2012), obtained from the EEA website, was used to assess the change in pollutants' concentration with increasing altitude. In general, locations at higher altitudes are less populated; the pollution dispersion processes are also easier to occur (Hu et al., 2014).

The land surface temperature (LST, 2016) and the normalized difference vegetation index (NDVI, 2016) generated from the MODIS Aqua and Terra platforms, the night time lights (NTL, 2012) product from the National Oceanic and Atmospheric Administration (NOAA), as well as the climatic data (humidity, precipitation and wind speed) from the National Centers for Environmental Prediction (NCEP) Climate Forecast System (CFSv2) (Saha et al., 2011) were pre-processed using the Google Earth Engine (GEE) API (Google Earth Engine Team, 2015). GEE makes it possible to rapidly process vast amount of satellite imagery on global