Atmospheric Environment 112 (2015) 234–245

Contents lists available at ScienceDirect

Atmospheric Environment

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

Comparing apples with apples: Using spatially distributed time series of monitoring data for model evaluation

E. Solazzo, S. Galmarini*

European Commission, Joint Research Centre, Institute for Environment and Sustainability, Air and Climate Unit, Ispra, Italy

ARTICLE INFO

Article history:
Received 3 February 2015
Received in revised form
9 March 2015
Accepted 15 April 2015
Available online 17 April 2015

Keywords:
Model evaluation
Spatial representativeness
Hierarchical clustering
KZ-filter
AQMEII

ABSTRACT

A more sensible use of monitoring data for the evaluation and development of regional-scale atmospheric models is proposed. The motivation stems from observing current practices in this realm where the quality of monitoring data is seldom questioned and model-to-data deviation is uniquely attributed to model deficiency. Efforts are spent to quantify the uncertainty intrinsic to the measurement process, but aspects connected to model evaluation and development have recently emerged that remain obscure, such as the spatial representativeness and the homogeneity of signals subjects of our investigation. By using time series of hourly records of ozone for a whole year (2006) collected by the European AirBase network the area of representativeness is firstly analysed showing, for similar class of stations (urban, suburban, rural), large heterogeneity and high sensitivity to the density of the network and to the noise of the signal, suggesting the mere station classification to be not a suitable candidate to help select the pool of stations used in model evaluation. Therefore a novel, more robust technique is developed based on the spatial properties of the associativity of the spectral components of the ozone time series, in an attempt to determine the level of homogeneity. The spatial structure of the associativity among stations is informative of the spatial representativeness of that specific component and automatically tells about spatial anisotropy. Time series of ozone data from North American networks have also been analysed to support the methodology. We find that the low energy components (especially the intra-day signal) suffer from a too strong influence of country-level network set-up in Europe, and different networks in North America, showing spatial heterogeneity exactly at the administrative border that separates countries in Europe and at areas separating different networks in North America. For model evaluation purposes these elements should be treated as purely stochastic and discarded, while retaining the portion of the signal useful to the evaluation process. Trans-boundary discontinuity of the intra-day signal along with cross-network grouping has been found to be predominant. Skills of fifteen regional chemical-transport modelling systems have been assessed in light of this result, finding an improved accuracy of up to 5% when the intra-day signal is removed with respect to the case where all components are analysed.

1. Introduction

The use of monitoring data for modelling applications (model development, data assimilation, and model evaluation) is widely documented. In a recent cooperative effort, the AQMEII (Air Quality Model Evaluation International Initiative) community (Rao et al., 2011) exploited monitoring air quality data for regional-scale model evaluation. Model outputs and observations from ground level monitoring networks were paired in time and space to quantify model performance and skill. As it is customary in model evaluation studies (AQMEII as well as many others), the deviation between air quality (AQ) models and measurements was solely attributed to model deficiency, neither the quality of the measurements nor the representativity of the monitoring sites and how different those can be in nature from a modelling result, are ever questioned. The latter points are the driving consideration motivating this study. The research questions we pose are:

- Leaving aside the instrumental uncertainty (e.g., Denby et al., 2011), is the nature of observational data really comparable to modelling results?

* Corresponding author.
E-mail address: stefano.galmarini@jrc.ec.europa.eu (S. Galmarini).

http://dx.doi.org/10.1016/j.atmosenv.2015.04.037

- Can we estimate cross-network consistency and spatial representativeness of observational networks prior to comparing monitoring time series to model results?
- Assuming that monitoring data and models have drastically different statistical supports, what is the level of intersection of their representativity that would make a comparison meaningful?
- What impact does the knowledge of that information have on model evaluation?

Typically, the preparation and use of observational data for model evaluation and development consists of:

1. Identifying monitoring stations/networks based on model grid size to make sure that the scale of the modelled phenomena we want to evaluate (e.g. regional or local-scale) is reflected in the observational data.
2. Select those stations that uniformly cover the whole simulation period without extended missing periods;
3. Space and time averaging of observational data is a common strategy to simplify the analysis.

For the evaluation of a regional scale AQ model, with a grid spacing of 10 km or more (i.e. the side length of the cell), roadside measurements would not be relevant since they capture features that would not be explicitly represented in these models. This aspect stringently impinges onto the representativeness of the measuring station (Larssen et al., 1999; Hogrefe et al., 2014; Janis and Robeson, 2004; Gego et al., 2005), which should be comparable with the grid spacing of the model or the scale of the processes that the model explicitly resolves (model resolution). The problem of matching a point measurement on the Earth surface with the volume-average for the layer of atmosphere closest to the surface over the extent of the model grid cell is known as 'incommensurability' (Swall and Foley, 2009).

Step (1) above is normally based on the classification of stations provided by the network of origin, normally complying with the internationally adopted classification convention. In the European Union (EU) the classification of AQ monitoring is that proposed by the European Environment Agency (EEA) and the European Air quality database (AirBase, more in Section 2) where stations are classified into urban, rural and sub-urban. Current practice in evaluation of regional models is to include only stations classified as rural which should, in theory, be representative of larger areas (Im et al., 2014a; Solazzo et al., 2012a; van Loon et al., 2007; Vautard et al., 2009). Station categories are assigned based on subjective criteria (at least in the European context), leading to non-homogeneous information and which do not always allow capturing subtleties in the station characterization. Henne et al. (2010) suggested to extend the current three AirBase categories to six, based on cluster analysis of ozone and NO_2 data. In the context of air quality spatial time series of data objective classification of air quality monitoring time series by Joly and Peuch (2012) set to eight the number of indicators needed to distribute pollutant time series into ten classes that were found to be pollutant-specific. Spangl et al. (2007) warned, however, about the inclusion of too many parameters in the site categorisation which might lead to superfluous sub-groups and difficult data interpretation. The extension of sub-regions over which models are evaluated is typically selected based on a-priori assumptions of, e.g., homogeneity of emissions or influence of the local orography (Solazzo et al., 2012a,b; Im et al., 2014a,b). The choice is then subjective, resulting in areas of mixed conditions and prevent the comparability among studies relying on different sub-regions.

After the description of the dataset (Section 2), in this study we introduce a method for the estimation of the area of representativeness of monitoring receptors (Section 3) and the limitations it imposes on its use for model evaluation. We then propose a novel methodology consisting in studying the associativity of the spectral decomposition of the pollutant time series rather than the raw data (Section 4). The criterion beyond such choice consists in assuming that components pertaining to different scales may show different levels of associativity and still be usable selectively to evaluate their counterpart in model data. Should that be the case only the components of similar associativity should be compared with the modelled counterpart. Conclusions are drawn in Section 5.

2. The data used

In this study we analyse the hourly ozone records for the year of 2006 used in the first phase of the AQMEII activity. Data for EU were derived from the AirBase (http://www.eea.europa.eu/themes/air/airbase) for a total of 1496 stations. AirBase is a public database containing air quality data from national monitoring programs from over than 30 participating European countries. It is managed by the European Topic Centre on Air and Climate Change (ETC/ACC) on behalf of the EEA. The sites are categorised in terms of station-type (traffic, industrial, residential, back-ground) and of area-type (urban, sub-urban, rural) based on population-density used as proxy of pollutant emission. AirBase data are widely used to support model evaluation studies other than AQMEII as part of the national and EU level operation air quality monitoring and reporting. Ozone datasets in North America (NA) have been also used to support the robustness of the methodology and the interpretation of the results. Ozone data for NA were prepared from hourly data collected by the AIRS (Aerometric Information Retrieval Systems, http://www.epa.gov/air/data/aqsdb.html) and CASTNet (Clean Air Status and Trends Network, http://java.epa.gov/castnet/) networks in the United States and the NAPS (National Air Pollution Surveillance, http://www.ec.gc.ca/rnspa-naps/) network in Canada. For full details on the data and the use made in AQMEII refer to Solazzo et al. (2012a,b,c, 2013b). Information about instrumental settings, sensitivity, and data acquisition protocols of each network can be retrieved on the internet page of the competent agency.

As quality check we only included in the analysis station whose valid record was higher than 85%, resulting in the removal of ~15% of the initial stations. Moreover, stations with more than 360 continuous hourly (15 days) concentration missing data were also removed from the analysis. Finally, missing records shorter than 360 h were interpolated using simple iterative linear regression imputation to facilitate the correlation analysis presented later (the data imputation has no impact on the results). Similar procedure was adopted in AQMEII for selecting the stations to be used for model evaluation (Solazzo et al., 2012a,b; Im et al., 2014a,b).

3. Spatial representativeness

The area of representativeness (AoR) of a monitoring station is typically defined as the spatial extension of the well-mixed air in which concentration of a given pollutant is homogeneous down to a given threshold (Larssen et al., 1999). Such a definition is applicable to single time averaged values. For time series the definition above is more practically extended to correlation of time series: AoR is *the area around a station in which other stations exhibit similarity above a chosen cut-off* (Orlowsky and Seneviratne, 2014), and relies on the density of the network. Because of the impossibility of knowing the concentration at each point around the receptor (or the time series), AoR is in practice estimated using several methods: proxy variables (Solazzo et al., 2013a; Martin et al., 2014); back-trajectory models (Henne et al., 2010); variogram analysis