

Interactive comment on “Strategies of Method Selection for Fine Scale PM_{2.5} Mapping in Intra-Urban Area Under Crowdsourcing Monitoring” by Shan Xu et al.

Anonymous Referee #1

Received and published: 5 February 2019

Xu et al describe measurements and spatial modeling of PM_{2.5}. Measurements were conducted with hand-held optical particle monitors. The spatial modeling compared multiple methods: ordinary kriging, universal kriging, and land use regression.

The paper suffers from several critical flaws and is not publishable in its current form. Below I outline five major problems with the manuscript.

Major Issue #1: I do not know what the authors mean by a "crowdsourced" data collection. The authors seem to define crowdsourcing in lines 27-28 of page 2, but "Crowdsourcing activities based on informal social networks and web 2.0 technologies that allowed citizens themselves to produce geospatial data among others" seems more

C1

like corporate jargon than a useful explanation of crowdsourcing.

The sampling approach seems to be short-term saturation sampling - many volunteers simultaneously sampled at predetermined locations. This sampling approach does not fit my personal notion of crowdsourcing, which would be a more informal data collection leveraging people's normal movements throughout the day. Sending an army of students to collect data in an organized fashion seems less like "crowdsourcing" and more like a sampling campaign. In that sense, this study has little distinction from the large literature on distributed air quality sampling.

What would be the value or longer-term viability of this or a similar sampling approach? This paper focuses on two short sampling periods of a few hours each, so the data are unlikely representative of long-term spatial patterns. Do the authors expect to deploy an army of distributed samplers on a semi-regular basis in order to build up a dataset capable of reproducing longer-term trends? Or to send out volunteers daily to make daily maps? I don't see how the "crowdsourced" aspect of this adds value or novelty; instead it seems like crowdsourcing is being used as a buzzword.

Major Issue #2: Data quality. Figure 1 shows one short-term comparison between the handheld PM monitors and the regulatory monitors. While there is generally good agreement, there is a fair amount of scatter among the handheld monitors. This scatter is to be expected given the low cost and the use of optical particle detection. However, the authors do not address how uncertainty in the measurements potentially impacts the mapping. Nor do they seem to account for uncertainty in the measurements or make any efforts to correct the measurements (e.g., based on hygroscopic growth).

Table 3 and section 3.1 - the crowdsourced data read higher PM than the regulatory data. The authors have not convinced me that this is not an artifact of the sensors they have chosen. During some hours there is significant difference between the mean "crowdsourced" PM and the mean regulatory PM. Since the overall spatial extent of the two sampling domains (regulatory and crowdsourced) is roughly similar, I would expect

C2

similar mean concentrations from each dataset.

Line 30 on page 8 calls the national monitoring sites "inaccurate." I am not familiar with regulatory measurement policies in China, but if they are anything like the US and Europe, the accuracy standard is high. The spatial pattern derived from these few monitors may be erroneous, but the specific measurements are accurate.

Major Issue #3: Site selection and sampling strategy The description of the sampling strategy is insufficient. Were all samplers deployed simultaneously at all sites in Table 1? How were the sampling times defined and chosen? What are significant differences between period 1 and period 2?

Table 1 - A better description of each type of site is needed. For example, Dust surfaces seem to be defined as "dust surfaces," which is not helpful to readers. What qualifies as a dust surface? Some entries in this table have "A" and "U". What do those designations mean?

Major Issue #4: Modeling and interpretation. The modeling aspect of this paper is not novel. Since the sampling method seems to be a straightforward saturation sampling campaign, using the resulting data to build spatial models is not a novel contribution. Numerous papers have already done this for PM_{2.5}, as noted by the authors.

One main conclusion seems to be that the modeling approaches work. This is not all that novel - it is more a statistical finding than an atmospheric measurement technique. Numerous papers have shown that LUR and kriging models can be fit to spatially distributed measurements.

Another conclusion is that the models work better when provided with more training sites. Again this seems like an obvious outcome, especially for the kriging approaches.

A more relevant analysis would be to evaluate if the models (and measurements) make physical sense. In Figure 5 there is a PM hotspot in the northwestern part of the domain on Day 1 and in the center of the domain on Day 2. Do these hotspots make sense

C3

given the distribution of sources and the climatology?

Major Issue #5: The paper needs a thorough review and edit for English grammar. There are many grammar errors (too many to count or enumerate here), and in other places the language is hard to follow.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2018-402, 2019.

C4