Visual Analytic Design for Detecting Airborne Pollution Sources

VAST Challenge 2017 Award: Comprehensive Mini-Challenge 2 Answer

Jo Wood*

giCentre (http://gicentre.org), City, University of London

Abstract

Using the VAST Challenge 2017 dataset as an illustration, the design choices of a visual analytic system for predicting the source of air pollution is described. Probabilistic Source Cones are visual symbols representing the probability of source location of a pollution event. Using transparency to indicate probability, multiple cones may be overlaid in order to provide a fuzzy triangulation of likely sources. This enabled the correct prediction and elimination of pollution sources at a precision far in excess of the spatial density of the sensors themselves.

1 Introduction

This paper describes some of the design choices that were made in building a visual analytics system to detect spatio-temporal patterns in airborne pollutants. It is illustrated using the fictional data and scenario described in the VAST Challenge 2017 Mini Challenge 2 (`vacommunity.org/VAST+Challenge+2017`). The challenge involved identifying the most likely sources of four different pollutants based on visual analysis of nine fixed chemical sensors. Readings from each sensor were logged every 15 minutes over three month-long periods spanning April-December 2016. Additionally, wind direction and speed logged at three-hour intervals and the locations of four industrial sites, which may or may not have been emitting pollutants were provided. The challenge, involving geospatially located multivariate data that change over time, has much in common with tasks suited to analysis with Geographic Information Systems (GIS). But the inherent uncertainty in the data (both due to data error and complex processes of plume dispersion) make this task well suited to visual analytics capable of representing spatial uncertainty.

2 Sensor Data Error Detection and Correction

Sensor data were initially inspected for errors by representing pollutant levels over time as line charts with a vertical scaling of the square root of molecular concentration (to show variation in the non-peak concentrations). Missing and multiple readings for the same sensor/timestamp were automatically detected and symbolised. The vertical alignment of the line charts revealed systematic gaps in the data as well as misclassification from one of the sensors. These were corrected before continuing with the analysis.

Statistical Process Control Charts [1] and CUSUM charts (as used in the VAST Challenge 2016) were used to detect suspicious events (e.g. structure in times of high concentration pollution releases) and trends in detected pollution levels. Figure 1 shows the CUSUM chart for a failing sensor that depicts a linear increase in pollutant concentrations over time (a random background level of pollutants would be represented as horizontal, not rising, CUSUM lines). While this sensor showed an anomalous drift not present in other nearby sensors, the distribution of peak concentration events (as defined by a user-selectable number of standard deviations above mean background levels) was consistent with other sensor readings.

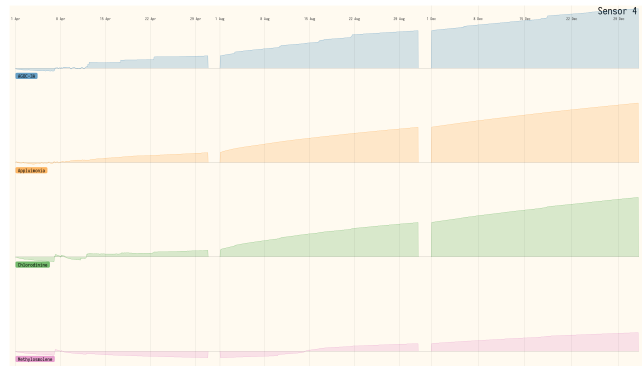

Figure 1: Drift in Sensor 4 readings revealed in CUSUM chart. Here the cumulative deviation from expected readings (based on the first week of April) is shown. All 4 chemical readings show an apparent gradual increase in the underlying concentration in addition a day-to-day noisy variation and irregular spikes.

3 Spatial Analyis

The location of eight of the nine sensors formed most of a ring surrounding four industrial sites that were candidates for potential pollution sources. A visual analytics system was designed to provide a map view of the sensors and chemical detection events. In this way, by considering the spatial intersection of multiple detection events, sources of pollution could be triangulated. This triangulation process was confounded by the fact that the sensors were some distance (typically 2-5km) from their source and that atmospheric conditions changed over the nine month period of observations. This meant there was considerable uncertainty about the spatial origin of any source of pollutants detected by the sensors.

As with some previous VAST challenges, analysis of these spatio-temporal patterns using fixed sensor data falls into a category of problem common in handling distributed geospatial networks [2]. That is, the desired spatial and temporal resolution for analysis is finer than that provided by the fixed sensor network. Nevertheless, symbolising the extent of spatial uncertainty in predicted plume location proved a successful strategy.

3.1 Probabilistic Source Cones

Every pollution event, identified by a detected chemical concentration of at least 5 standard deviations above background concentrations, was symbolised with a *probabilistic source cone (PSC)*; see Figure 3. The cone originates from the detecting sensor and is orientated in the direction of the wind at the time of detection event. The length of the PSC is proportional to the strength of the wind at that time. The cone itself is rendered using transparency, with increasing transparency at 5 degree intervals away from the monitored wind direction (which has a low spatial and temporal precision). Thus a PSC provides a visual clue about the origin of a detected event.

*e-mail: j.d.wood@city.ac.uk

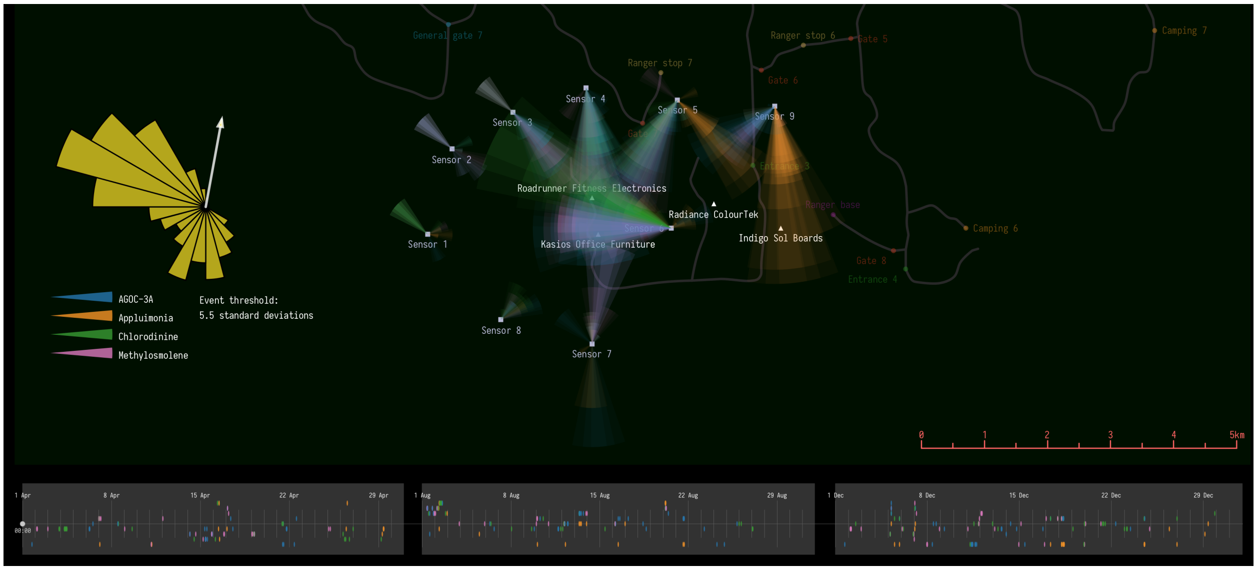

Figure 2: Composite of all chemical detection events of at least 5.5 standard deviations from background levels. Probabilistic Source Cones (PSCs) show likely origin of four chemicals types.

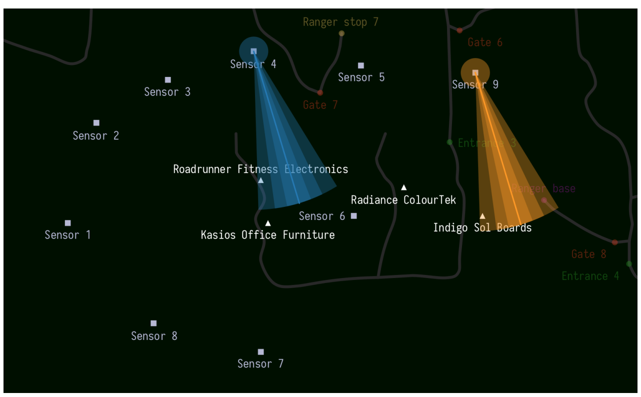

Figure 3: Two simultaneous pollution detection events symbolised as a probabilistic source cone. The direction of the cone is based on the closest monitored wind direction to the detection event and the length of the cone proportional to wind speed.

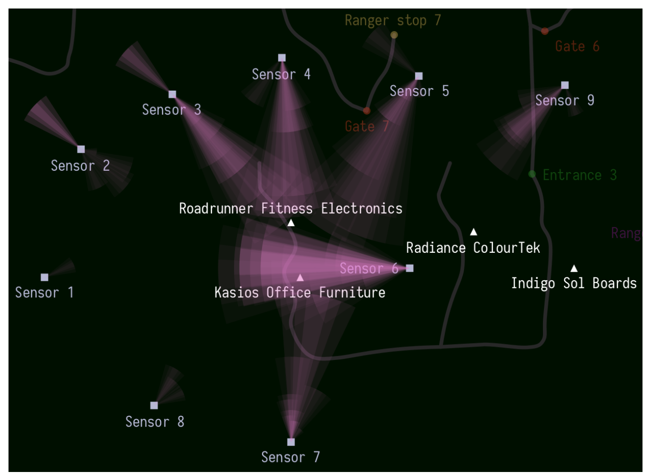

Figure 4: Methylosmolene PSCs for pollution events of at least 5 standard deviations from background levels..

In isolation, a single PSC provides only a limited amount of evidence of a pollution source. But combined with multiple events, overlaid PSCs can reveal more insightful patterns. Figure 2 shows all pollution events at least 5.5 standard deviations above background. The intersection of PSCs provides a 'fuzzy triangulation' with a precision in excess of the spatial density of the sensor network. For example, Sensors 5 and 9 in combination suggest the source of the fictitious chemical Appluimonia (in orange) is the 'Indigo Sol Boards' factory.

Isolating the PSCs for each chemical type can show not only where sources appear to intersect, but also eliminate sources outside of the intersecting regions. For example Figure 4 shows combined PSCs for the fictitious chemical Methylosmolene where the intersecting cones suggest 'Kasios Office Furniture' and not 'RoadRunner Fitness Electronics' is the most likely source. If sensors 3, 4 or 5 had been considered in isolation, they would all have incorrectly suggested RoadRunner as the origin of the pollutant.

4 Conclusion

GIS tend to be best suited to spatial analysis of crisp objects with known boundaries. For tasks such as those in the 2017 VAST challenge, visual analytics that can represent uncertainty offer potential for greater insight. Visual Analytics using PSCs is suited to low spatial and temporal precision in wind monitoring. There is an arbitrary mapping of cone size, both length and angular deviation, to wind conditions, although interactive scaling of the cone dimensions allowed calibration of this mapping. For more precise readings, non-linear 'cones' could be derived where their orientation varies over their length.

References

[1] S. Bersimis, S. Psarakis, and J. Panaretos. Multivariate statistical process control charts: an overview. *Quality and Reliability Engineering International*, 23(5):517–543, 2007.

[2] M. Duckham. *Decentralized spatial computing: foundations of geosensor networks*. Springer Science & Business Media, 2012.