MC2 - Mining Factory Pollution Data through a Spatial-Nonspatial Flow Approach (Honorable Mention for Clarity in Visual Communication)

Joshua Castor* Joseph Borowicz* Andrew Burks*
Manu Thomas* Timothy Luciani* G.Elisabeta Marai*

Electronic Visualization Laboratory, University of Illinois at Chicago

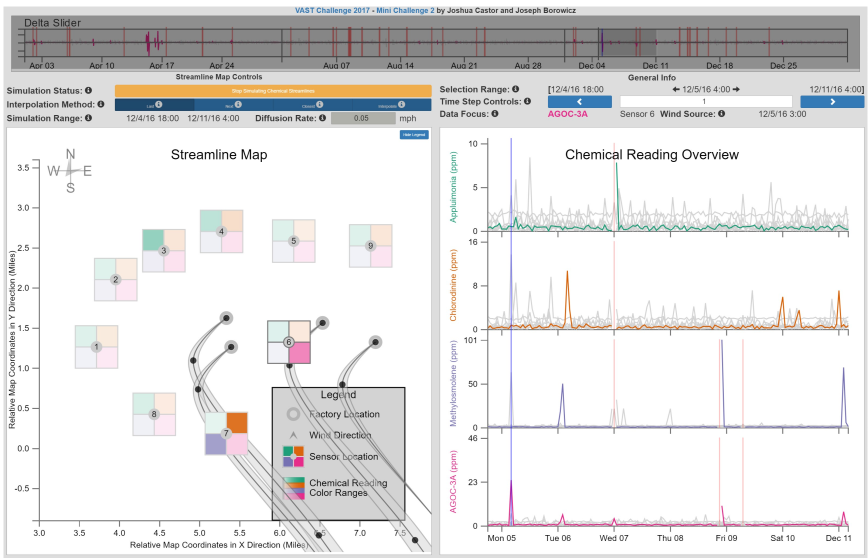

Figure 1: Interactive pollution data analyzer. The raw data at the current time stamp shows the wind is moving southwest, but the simulation lines show that the movement of chemicals from factories is on a curved path. Overlapping path lines with spikes in chemical readings (indicated by brighter colors) reveal that the sources of said spikes may come from the three factories.

Abstract

Mini Challenge 2 of the VAST Challenge 2017 focused on a small industrial area south of the fictional Mistford preserve, specifically around four manufacturing factories. Our main goal was to develop a visual analytics tool to explore the spatio-temporal chemical readings and wind data. Specifically, we wanted to determine which factories were responsible for emitting which chemicals and to determine the performance of the nine sensors in the area. In order to help achieve this goal, we developed a web-based application that utilizes interactive visualizations and path line analysis for revealing sensor errors and chemical reading spikes, along with pinpointing the possible sources of chemical reading spikes.

1 Introduction

Our solution is a web-based visual analysis tool, with linked views (Fig. 1). There is an overview calendar timeline at the top that displays the delta values of chemical readings. The left panel contains a map with the physical location of factories (circles) and sensors (square glyphs), and optional path lines that reconstruct particle pathways over time. A constant diffusion factor is added to the resulting pathways to show areas of possible emissions. A detailed chemical chart view is shown on the right, with the selected sensor data shown in color, and the other sensor data in gray. There are additional filters that control the path line simulation and the information shown.

Colors are mapped to the 4 different chemicals. Brushing and linking through color and interaction are used to correlate the views.

*{jcasto3, jborow5, aburks3, mthoma52, tlucia2, gmarai}@uic.edu

2 Methods

The data given with the challenge includes two datasets, sensor readings and wind readings, over the course of the three discrete months of April, August, and December 2016. The sensor readings are given every hour, per sensor, per chemical. The wind data is given as a single reading for the whole area, every three hours.

Path Line Simulations The path line simulation (that we created to analyze this data) starts with one seed point at each factory. For each timestamp simulated, a wind vector is added to the calculations, transposing the previous points to create a line. We include four wind vector options, because the wind data is only available for nearly every third hour. The options include using the last available wind data point, the next available data point, the closest available data point, and an interpolation between closest available data points. We add wind direction to the magnitude to create the offset vectors. A constant diffusion factor is added to the resulting path to show an area of possible emissions.

Through the path line approach, the simulation data results in patterns that are not easy to perceive from just a set of readings (Fig. 2). These path line patterns help to correlate wind movements with chemical reading spikes, giving an idea of possible sources of

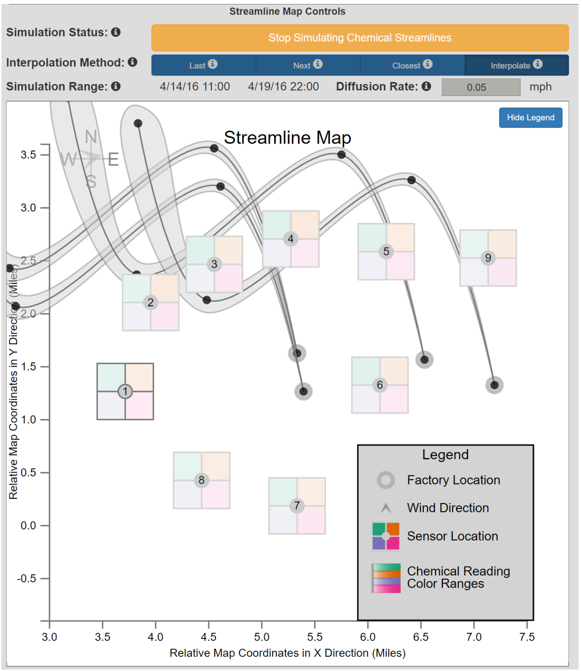

Figure 2: The wind data simulation reveals unusual and sometimes interfering path shapes.

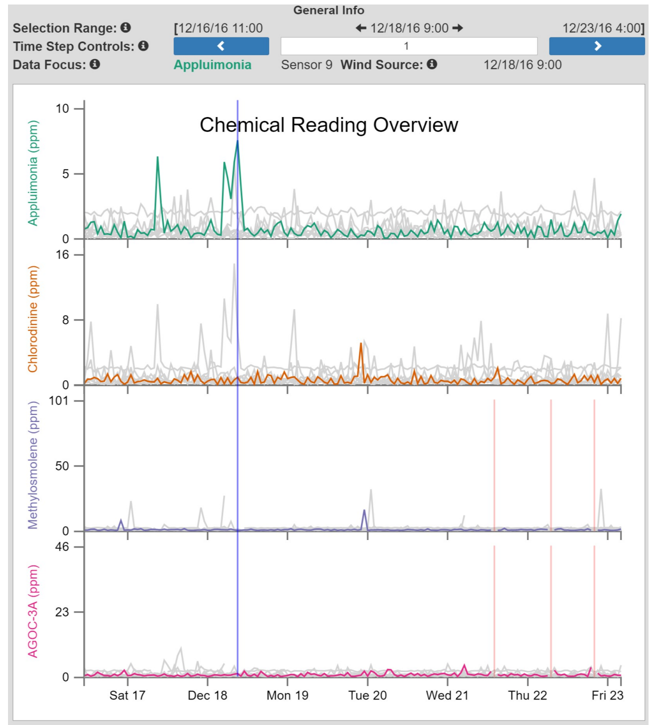

Figure 3: The chemical reading overview panel shows chemical spikes and drops throughout the chosen range for the selected sensor, as well as any sensor errors. The data from non-selected sensors is shown in gray. The blue vertical marker represents the current time stamp shown on the map view.

high readings (Fig. 1).

3 Results

3.1 Sensor Performance

Our analysis of wind data reveals that the wind sensor is not fully performant, as there is missing data. For example, we found a time period from Aug 1st 2016 0:00 to Aug 4th 2016 18:00 where we had no wind data (skipped data, not 0), indicating that the wind sensor was not functioning properly in that period.

The sensor readings had multiple errors, mostly missing data (Fig. 3), which indicates that the sensors weren't always fully functional either. There are multiple instances, including the 2nd and 7th of every given month, where we were missing chemical readings (partially or completely) for nearly all of the sensors. Also in multiple cases, when Methylosmolene has no reading, AGOC-3A has 2 readings. For example, at timestamp 8/8/16 10:00, Methylosmolene has a null reading, which indicates that there's no entry for it, while AGOC-3A has two readings. What's interesting to note is that sometimes there will be a large difference in the two values for AGOC-3A.

3.2 Chemicals

Chlorodinine Roadrunner Fitness Electronics (RFE) is the primary polluter of Chlorodinine. This is shown by a path line simulation using wind interpolation starting from 12/22/16 16:00. This simulation shows peaks in sensor 6 at 12/23/16 0:00 and 12/23/16 5:00, and in both these cases the path line from RFE intersected sensor 6. Similarly, we see path line simulations starting from 8/11/16 10:00 and stopping the time slider on 8/11/16 22:00, 8/12/16 2:00, and 8/15/16 10:00 that capture similar behavior.

Methylosmolene Methylosmolene's primary polluter is Kasios Office Furniture (KOF). On 4/2 4:00 the path line emanating from KOF intersects sensor 6 showing a concentration in excess of 88 ppm. At 4/3 0:00 the path line intersects sensor 6 showing a concentration of 42 ppm. Similarly, the path line simulation at 4/9 1:00 clearly intersects the sensor 6 showing a concentration in excess of 94 parts per million of Methylosmolene.

AGOC-3A Radiance Colourtek (RC) is the primary polluter of the chemical AGOC-3A. Beginning 4/15/16 6:00 and extending to 4/15/16 12:00 the path lines emanating from RC that intersect sensor 6 show spike AGOC-3A concentrations that exceed 45 ppm. The AGOC-3A peaks on August 6th, 12th, and the 25th all have intersecting path lines emanating from RC as well. Finally, the largest spike in concentration of AGOC-3A in December occurred on the 15th at sensor 6 @4:00 also had a path line starting at RC intersecting at that point.

Appluimonia Indigo Sol Boards's (ISB) primary pollutant is Appluimonia. On 4/7 1:00 and 2:00, sensor 9 shows peaks of Appluimonia in excess of 4.9 ppm clearly has an interesting path line originating from ISB. This situation is repeated on August 10th. Similarly a path line originating at ISB on 12/15 intersecting sensor 9 during concentration peaks at 12:00 and 21:00.

4 Conclusion

In conclusion, we developed a web-based visual analytics tool that allows the capture of sensor anomalies and the study of correlations between wind patterns and chemical reading spikes. Our solution, which is implemented using D3 and javascript, blends spatial and nonspatial representations.

Acknowledgments

We thank the anonymous reviewers for their generous and supportive feedback. We thank everyone at the Electronic Visualization Laboratory for their support. This work was supported in part by NSF awards CNS-1456638 and IIS-0952720, and by the NSF GRF program.