Julius-Maximilians-
UNIVERSITÄT WÜRZBURG

Data Mining and Information Retrieval Group
University of Würzburg, Germany (1),
University of Kassel, Germany (2),
L3S Research Center, Hannover, Germany (3)

UNIVERSITY OF KASSEL
Knowledge & Data Engineering

Forschungszentrum · Research Center
L3S

# Subjective and Objective Data: Bridging the Gap

*Martin Becker [1,3], Andreas Hotho [1,3], Juergen Mueller [2,3], Mark Kibanov [2,3], Martin Atzmueller [2], Gerd Stumme [2,3]*

## EveryAware

Enhance environmental awareness
through social information technologies

SEVENTH FRAMEWORK PROGRAMME

http://cs.everyaware.eu

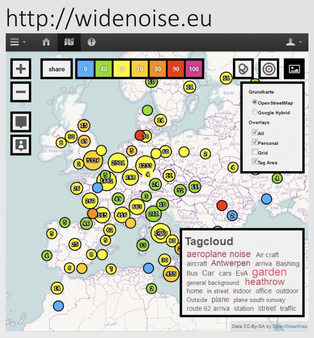
Partners

## WideNoise

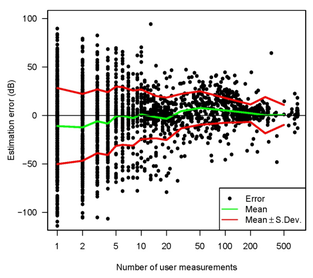
Collaborative noise pollution monitoring with smartphones

There are various kinds of pollution that often get on the first page of newspapers. Noise pollution on the other hand is rarely mentioned. Nevertheless, it is something that constantly surrounds us even though we do not notice. WideNoise helps users to better understand the soundscape around them and correlates objective noise measurements with subjective perceptions from users.


http://widenoise.eu

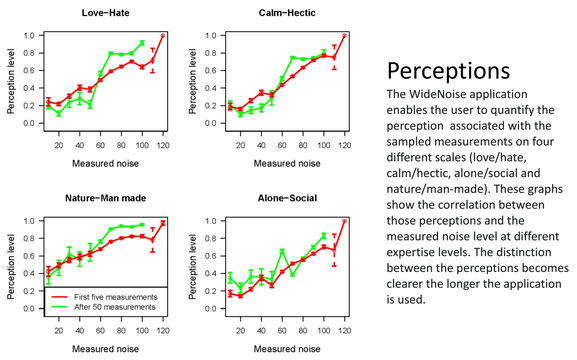



### Subjective Data




#### User Estimates

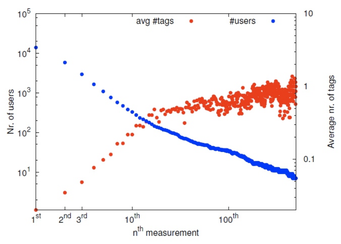
The WideNoise application allows users to estimate the noise reading before being provided with the real value. Each point in this plot corresponds to the deviation of the estimated noise reading from the real reading at different expertise levels. The longer the application is used, the better the estimation gets.




#### Perceptions

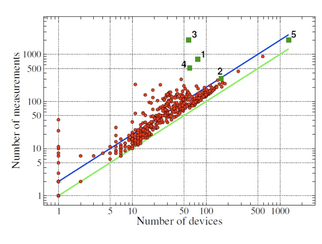
The WideNoise application enables the user to quantify the perception associated with the sampled measurements on four different scales (love/hate, calm/hectic, alone/social and nature/man-made). These graphs show the correlation between those perceptions and the measured noise level at different expertise levels. The distinction between the perceptions becomes clearer the longer the application is used.

### Dataset Statistics



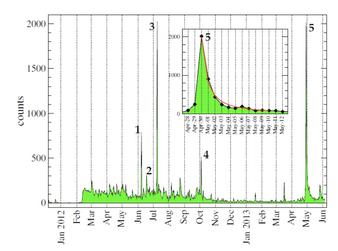

#### Tag Usage

This plot shows how often users attach tags in comparison to how often users measures noise. The cumulative number of users submitting at least *n* measurements is displayed in blue (left axis legend), while the red points represent the average number of tags used in the *n*-th users' measurement (right axis legend).




#### User Activity

This scatter plot shows the number of measurements collected each day compared to the number of active devices at that day. The dark green symbols correspond to the most important spikes/events mentioned in the noise measurement per day figure. The green and blue lines are guides for the eye and correspond to the case of one measure per device and two measures per device respectively.




#### Noise Measurements per Day

This plot illustrates the activity characteristics of WideNoise represented as the number of measurements collected each day from Dec. 8th 2011 till Jun. 6th 2013. The labels correspond to:

(1) a case study in Rome (9th June 2012);
(2) the launch of the Heathrow activities (19th June 2012);
(3) an Antwerp test case (10th July 2012);
(4) the Birmingham workshop (5th October 2012);
(5) an article in German regional newspaper (published April, 29th 2013, activity peak on April, 30th 2013).

In the inset an enlarged view of event 5 is shown. The decay of user participation is consistent with a power-law of exponent of approx - 1.33 (red curve).

| Location | Number of measurements | Area covered (km$^2$) | Total Time (hours) | Average noise level (dB) | Devices |
|---|---|---|---|---|---|
| World | 40661 (817 without location) | 4.80982 | 541.93 | 64.16 | 13962 |
| Europe | 27771 | 3.36757 | 354.8 | 65.98 | 7395 |
| Asia | 11033 | 1.1358 | 164.49 | 59.59 | 5392 |
| North America | 1373 | 0.232655 | 21.59 | 64.39 | 588 |
| South America | 93 | 0.015525 | 1.51 | 66.25 | 56 |
| Africa | 107 | 0.01597 | 1.70 | 67.42 | 47 |
| Australia | 193 | 0.02610 | 3 | 58.44 | 96 |

Becker, M.; Caminiti, S.; Fiorella, D.; Francis, L.; Gravino, P.; Haklay, M.; Hotho, A.; Loreto, V.; Mueller, J.; Ricchiuti, F.; Servedio, V. D. P.; Sîrbu, A. & Tria, F. (2013), 'Awareness and Learning in Participatory Noise Sensing', PLoS ONE 8 (12) , e81638.

Becker, M.; Caminiti, S.; Fiorella, D.; Francis, L.; Gravino, P.; Haklay, M.; Hotho, A.; Loreto, V.; Mueller, J.; Ricchiuti, F.; Servedio, V. D. P.; Sîrbu, A. & Tria, F. (2013), 'Awareness and Learning in Participatory Noise Sensing', PLoS ONE 8 (12) , e81638.

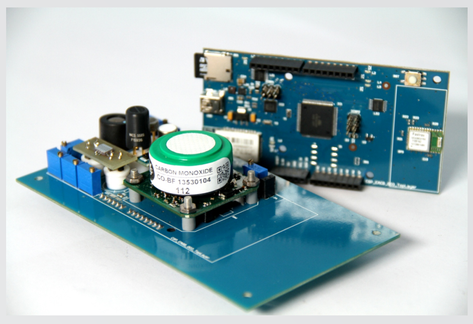
## AirProbe

Collective air quality monitoring

AirProbe is an air quality application developed to help citizens to be aware of the air quality around them independently of data provided by government agencies. The combined use of a sensor box, the AirProbe Android application and a web frontend lets the user collect, interpret and share the air quality in her personal environment. The tagging feature allows to add subjective impressions to the measured data.

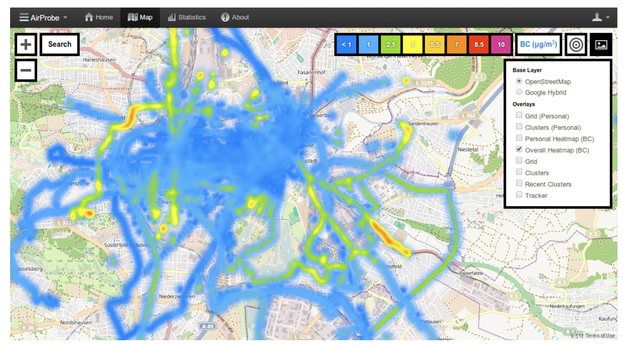



http://airprobe.eu




### Dataset Statistics

16,082,015 measurements
45 sensor boxes
46 users
50 devices
1,199 tags




#### Air Quality Heatmap

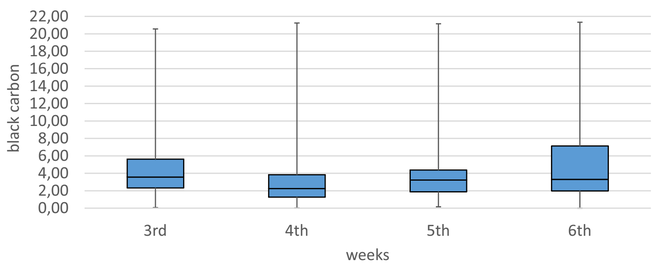
Heat map for Kassel, Germany, showing measured Black Carbon (BC) values

### The APIC Challenge

Within the EveryAware project we were challenging participants from four cities (Antwerp, London, Kassel and Rome) to cover the black carbon (BC) concentrations of their environment as completely as possible. The first phase was focused on a web game where users had to estimate the air quality in their city. Phase two and three (two weeks each) were about actually measuring BC concentrations with a mobile sensor box. The measured data was used as feedback in the web game.

#### Specific observations from the APIC challenge

(to be published in collaboration with our partners from the EveryAware project and especially Alina Sîrbu)




#### Development of measured black carbon (BC) concentrations

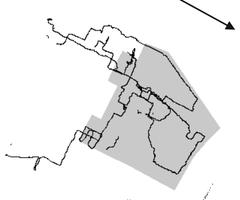
This plot shows the development of measured black carbon (BC) concentrations during the four weeks of the APIC challenge. Especially the last week shows a tendency for higher BC concentration, which can be explained by the participants specifically looking for sources of air pollution in their personal environment.

| Week of APIC | Measurements |
|---|---|
| 3rd | 1,409,770 |
| 4th | 1,548,607 |
| 5th | 1,364,537 |
| 6th | 1,990,658 |
| **Sum** | **6,313,572** |

#### Influence of coverage objectives

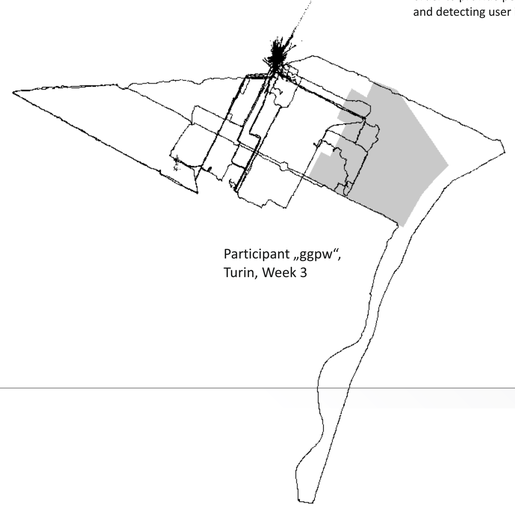
Coverage diagrams of a participant from Turin during four weeks of measuring in the context of the APIC challenge. In the first two weeks participants were asked to keep to a certain area. In the third and fourth week they were asked to measure where ever they wanted.

Focused exploration of a designated area will result in a better coverage and characterization of that area's air quality situation, as can be seen in the diagrams for the first two weeks. On the other hand, leaving the user free to choose the area to measure, will result in a more personal picture of her air quality environment, as can be seen in the bottom diagrams. Thus, in order to provide personalized feedback, analyzing, understanding and detecting user intention, is important.

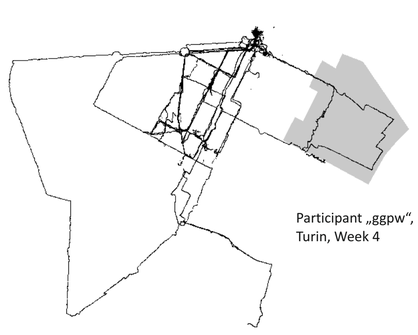


Participant „ggpw“, Turin, Week 1

Participant „ggpw“, Turin, Week 2

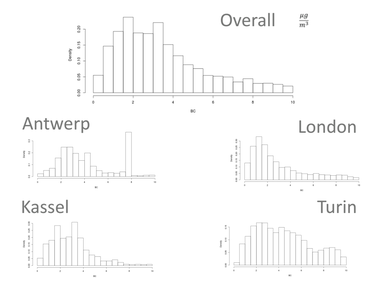


Participant „ggpw“, Turin, Week 3



Participant „ggpw“, Turin, Week 4

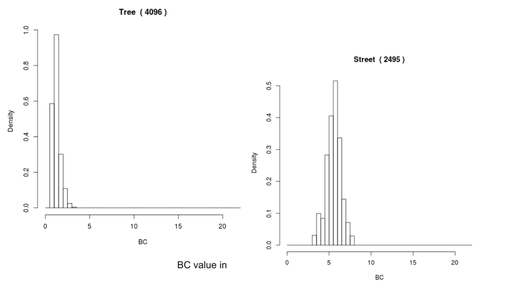
### City Statistics and Subjective Data




#### Measured Black Carbon Concentration

These plots shows the black carbon (BC) concentration distributions overall and in different cities. Different cities show different characteristics. Especially Turin has a rather bad footprint.

data collected between:
2013-04-11 00:00:00
2013-11-29 12:00:00




#### Black Carbon Concentration for Different Tags

These plots show black carbon (BC) concentration distributions for different tags. It is possible to distinguish and characterize different tags using these distributions.

## Work in progress

- **Modelling** the combined objective and subjective spatio-temporal data (using topic models)
- **Concept detection** by characterizing tags in different dimensions (e.g. by clustering or subgroup detection)
- **Tag recommendation** for spatio-temporal data in sparse environments (based on topic models)
- **Improving calibration** of the sensor box