ESA eo open science 2.0
12-14 October 2015, Frascati, Italy

Making sense of crowdsourced observations:

Data fusion techniques for real-time mapping of urban air quality

Philipp Schneider[1]
Nuria Castell[1]
William A. Lahoz[1]
and the entire CITI-SENSE team

[1]NILU – Norwegian Institute for Air Research


Air pollution
is the top
environmental
risk factor for
premature death
Getty


In 2011, 458000 premature deaths in Europe were attributed to particulate matter in the air
The Guardian

Traditional Air Quality Monitoring

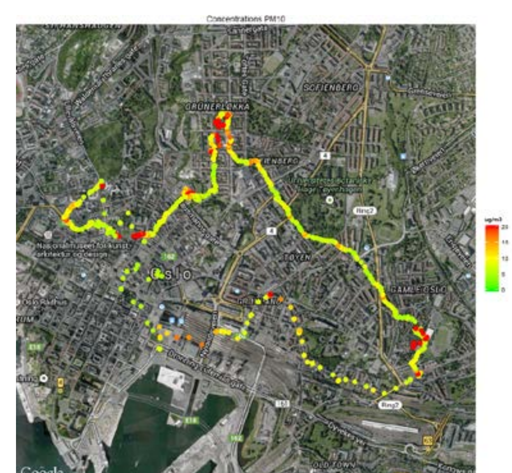
- Large
- Complex
- High-maintenance
- Expensive

→ Very sparse

Is there another way?

Bidgee

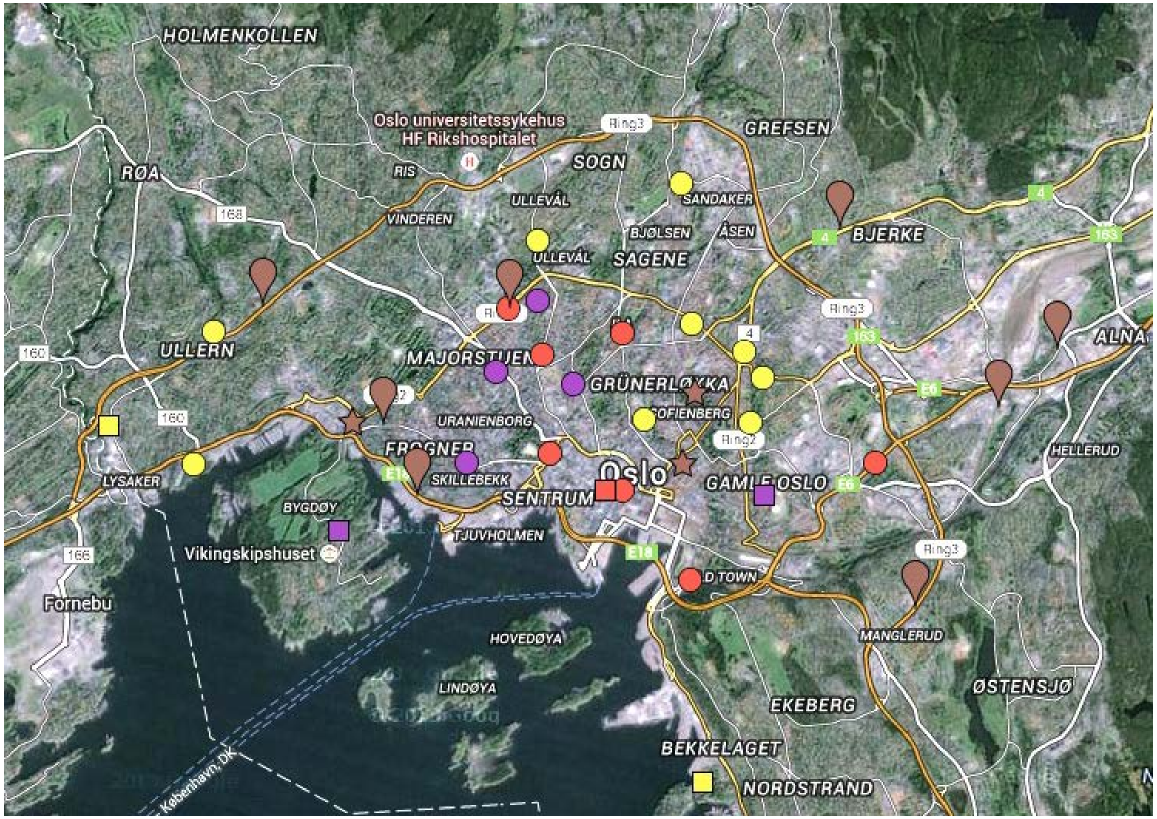
There might be…



Static observation network

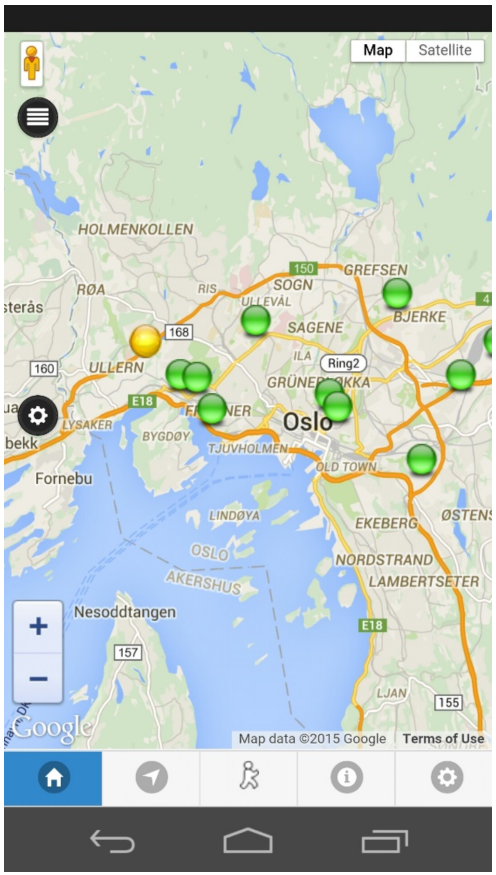
- AQMesh by Geotech
- Wireless air quality monitor
- Measures a variety of pollutants: NO, NO_2, O_3, CO, SO_2, PM_{10}, $PM_{2.5}$, as well as temperature, humidity, and pressure
- Compares reasonably well with reference equipment (but dependent on species)

Geotech
AQMESH



Schneider et al. (2015). Making sense of crowdsourced observations. ESA eo open science 2.0, 13 October 2015, Frascati, Italy

Point-based observations: The problem

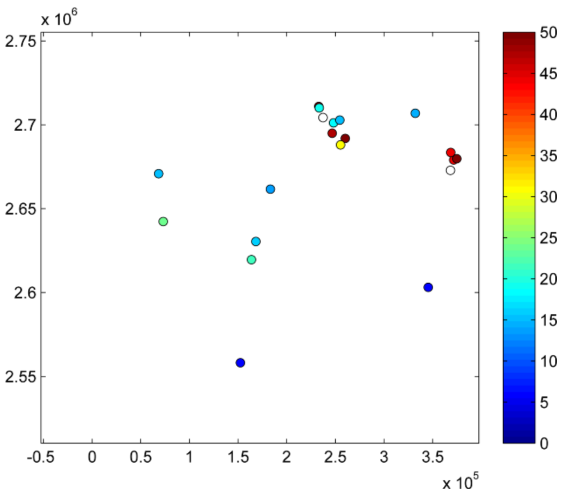


CITI-SENSE app: Android-based mobile app for real-time AQ monitoring

How do we get from this:

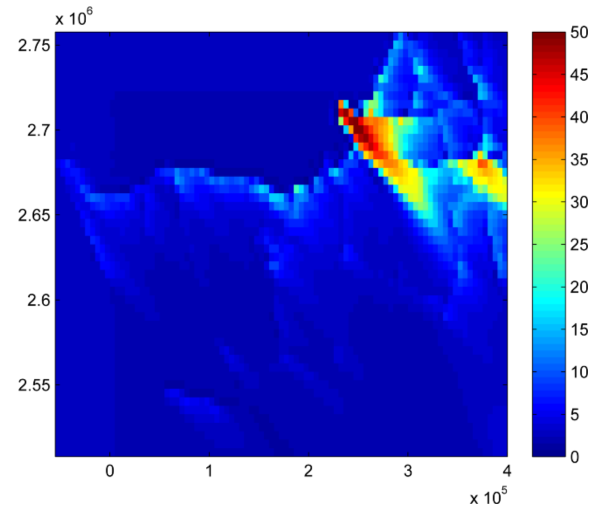
To this:

Data fusion: Basic Premise



Observations

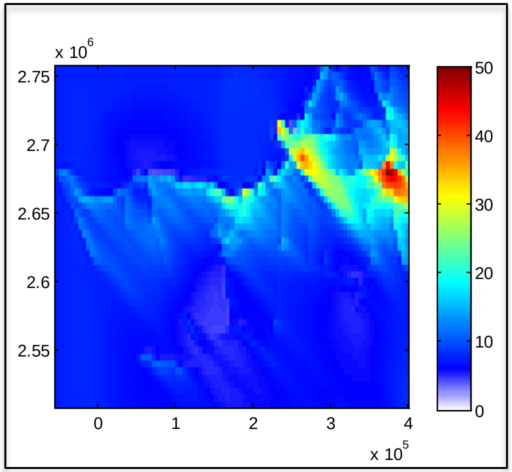
DATA FUSION



Modelling results or other auxiliary data

Data fusion (as a subset of data assimilation) creates a value-added product by

a) Interpolating the observations in an objective way
b) “correcting” the model estimates with true observations

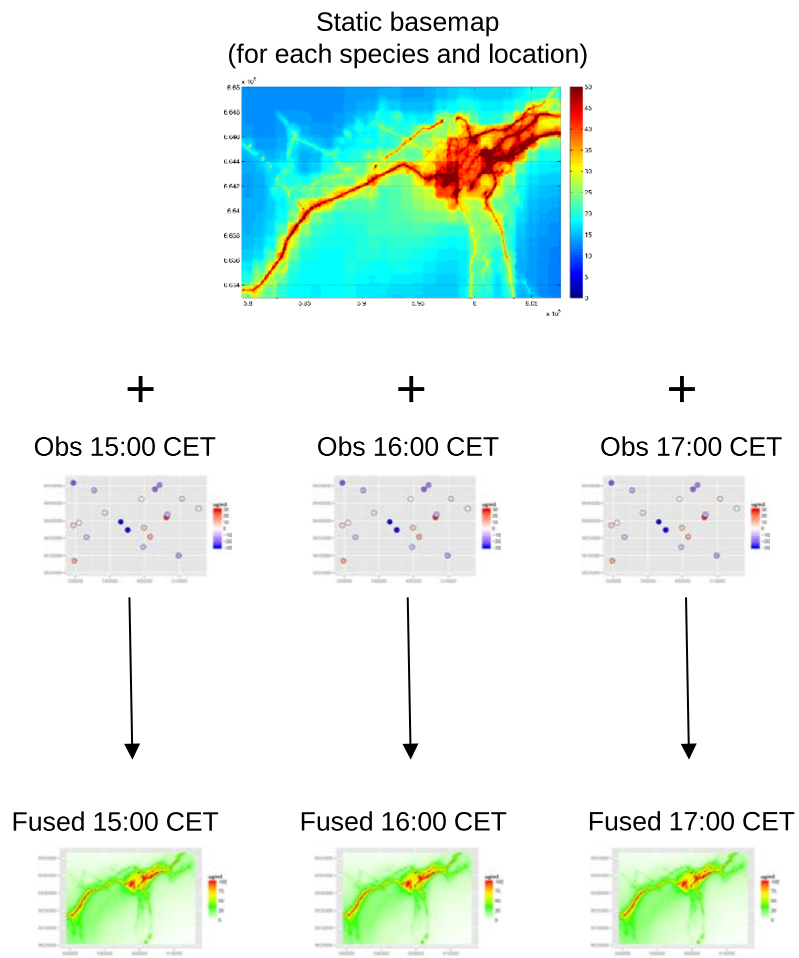


Combined map

Data fusion method used here provides a combined concentration field by regressing the observations against model data and spatially interpolating the residuals

Data fusion for CITI-SENSE

- A static basemap is created for each location and each species of interest to show the long-term spatial patterns
- This basemap is then modified according to the observations made by the static Geotech sensors
- This is essentially a location-dependent level-shift of the basemap
- The final result are hourly maps with the current best guess for the $NO_2/PM_{10}/PM_{2.5}$ concentration field at all CITI-SENSE locations



Basemap: Provides information about general spatial patterns

Geotech observations: Provide information about current state of atmosphere at a few sampling locations

Fused map: Value-added product providing a best guess of current state of atmosphere for the entire domain

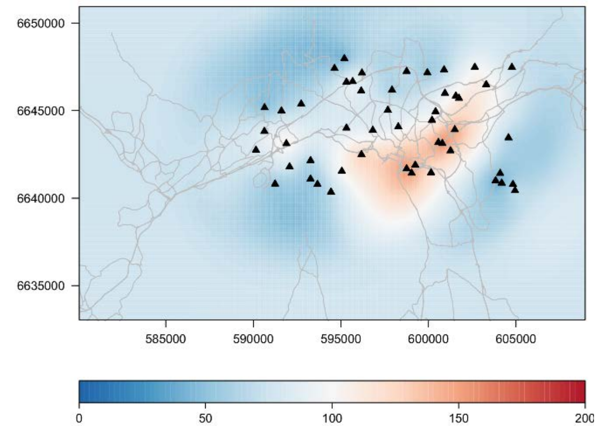
Schneider et al. (2015). Making sense of crowdsourced observations. ESA eo open science 2.0, 13 October 2015, Frascati, Italy

Data fusion methodology

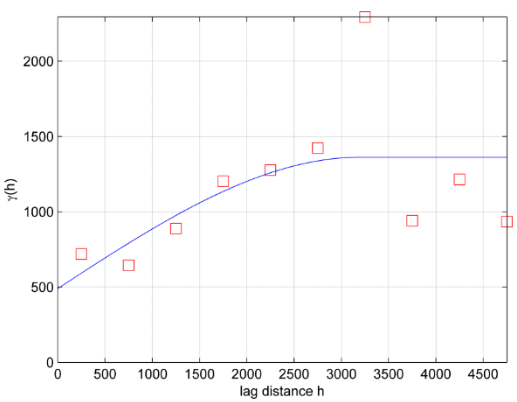
- Data fusion is a subset of data assimilation techniques (Lahoz and Schneider, 2014)
- Uses geostatistical framework
- Analysis performed entirely in log-space
- Universal kriging approach
- Spatial interpolation guided by proxy
- Explicit automated modelling of spatial autocorrelation

Lahoz, W. A., and P. Schneider (2014), Data assimilation: making sense of Earth Observation, *Front. Environ. Sci.*, *2*(16), 1–28, doi:10.3389/fenvs.2014.00016.

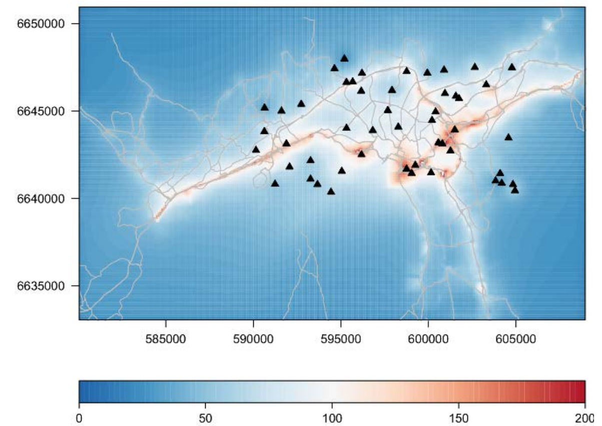
Using simple spatial interpolation



Theoretical model of spatial autocorrelation

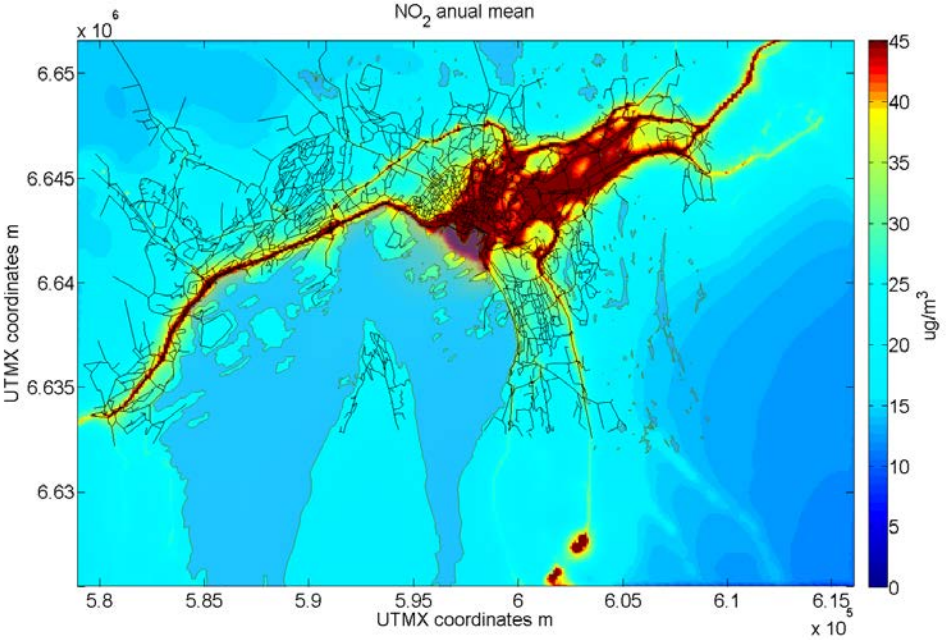


Using data fusion with spatial proxy

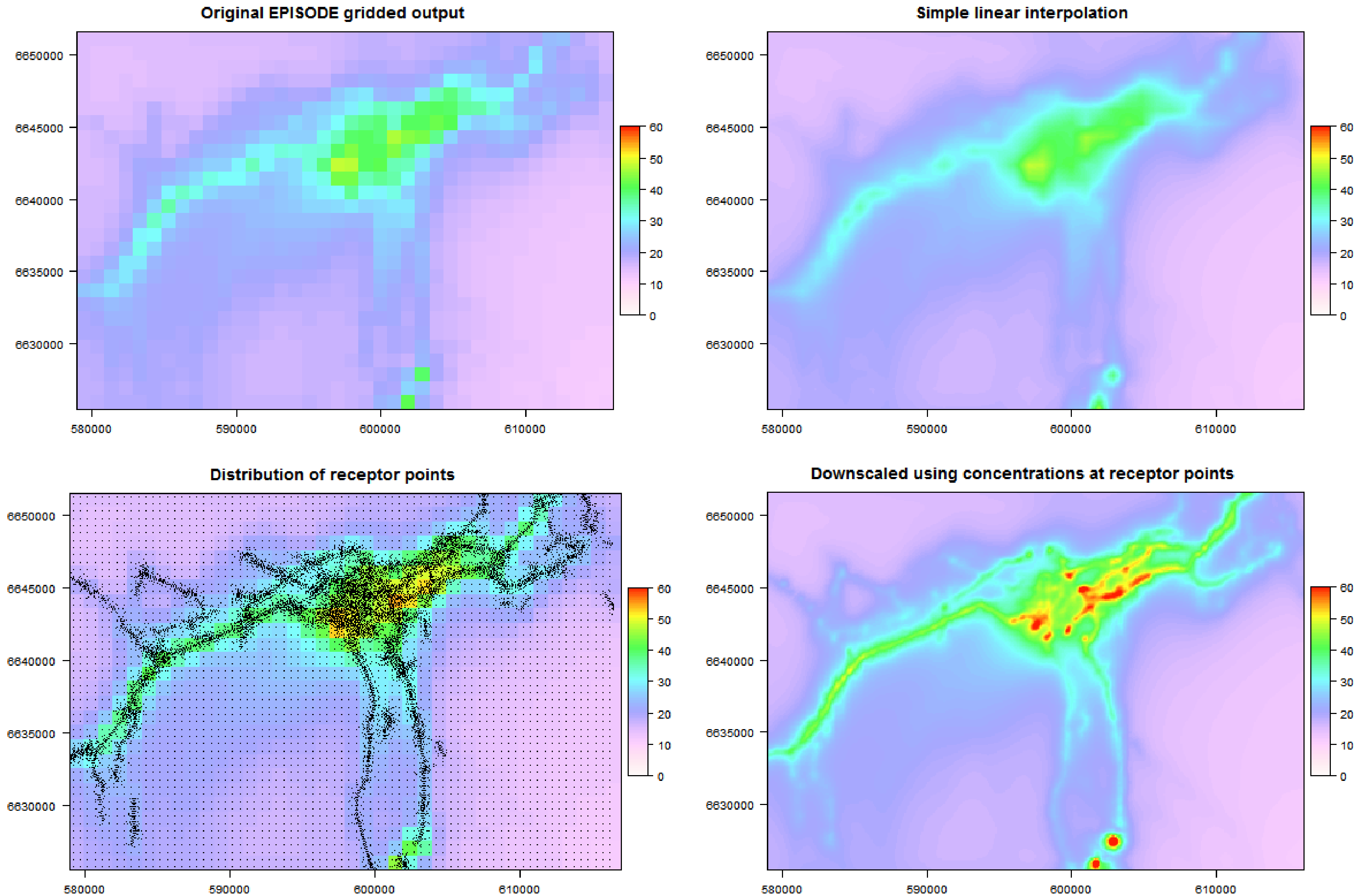


Modelling of the basemaps

- Can be nearly any spatially exhaustive dataset that is related to the observation
- Best to use are urban-scale dispersion models
- Alternatively concentration map created through LUR modelling
- We use the EPISODE model
  - Three-dimensional, combined Eulerian/Lagrangian air pollution dispersion model, developed at NILU
  - Combined modelling and postprocessing approach to obtain basemaps at 10-100 m spatial resolution

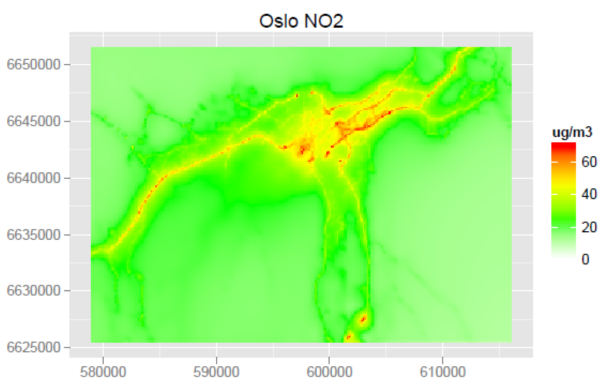


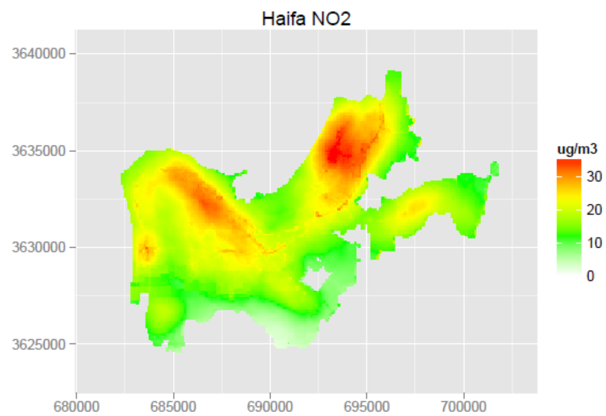
High-resolution map of NO_2 in Oslo from the EPISODE dispersion model. These kind of maps are ideally suited as a spatially distributed auxiliary dataset.

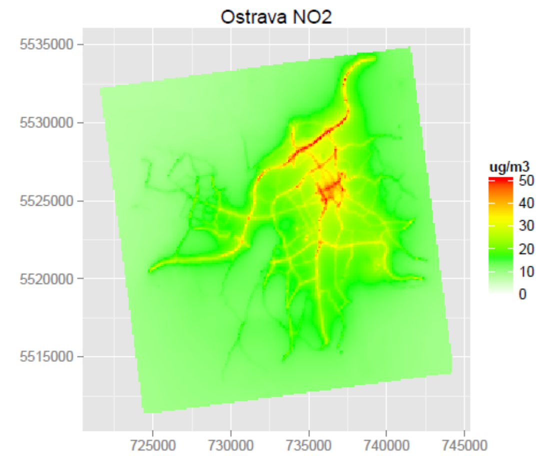


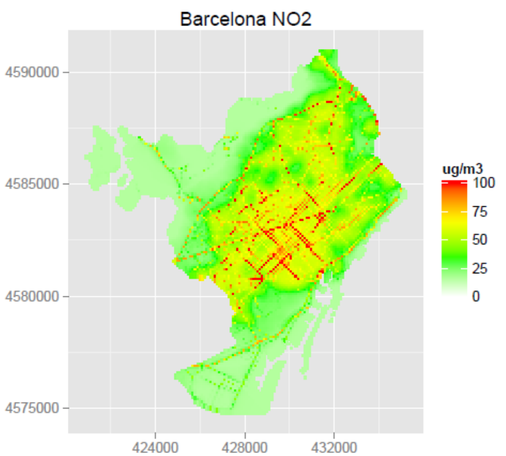
Receptor-point based downscaling of the gridded EPISODE output

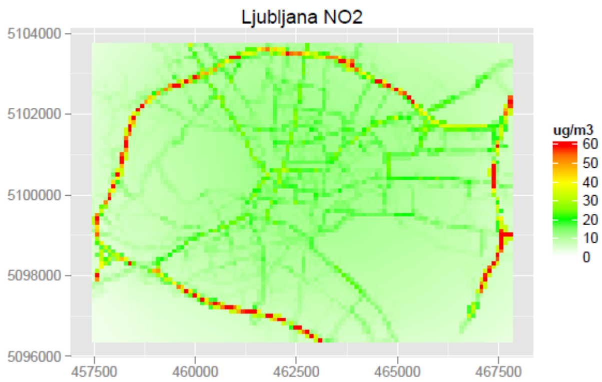
Example basemaps for NO_2

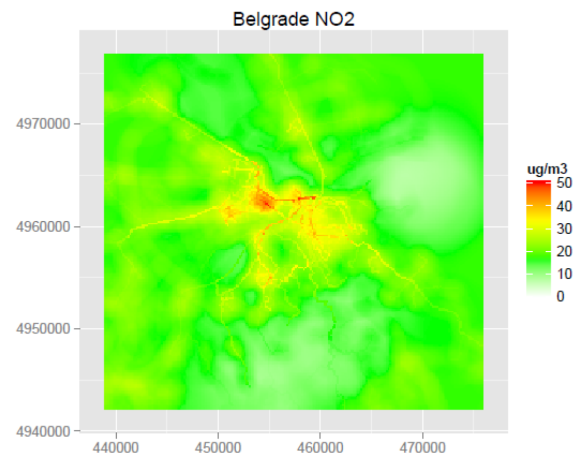






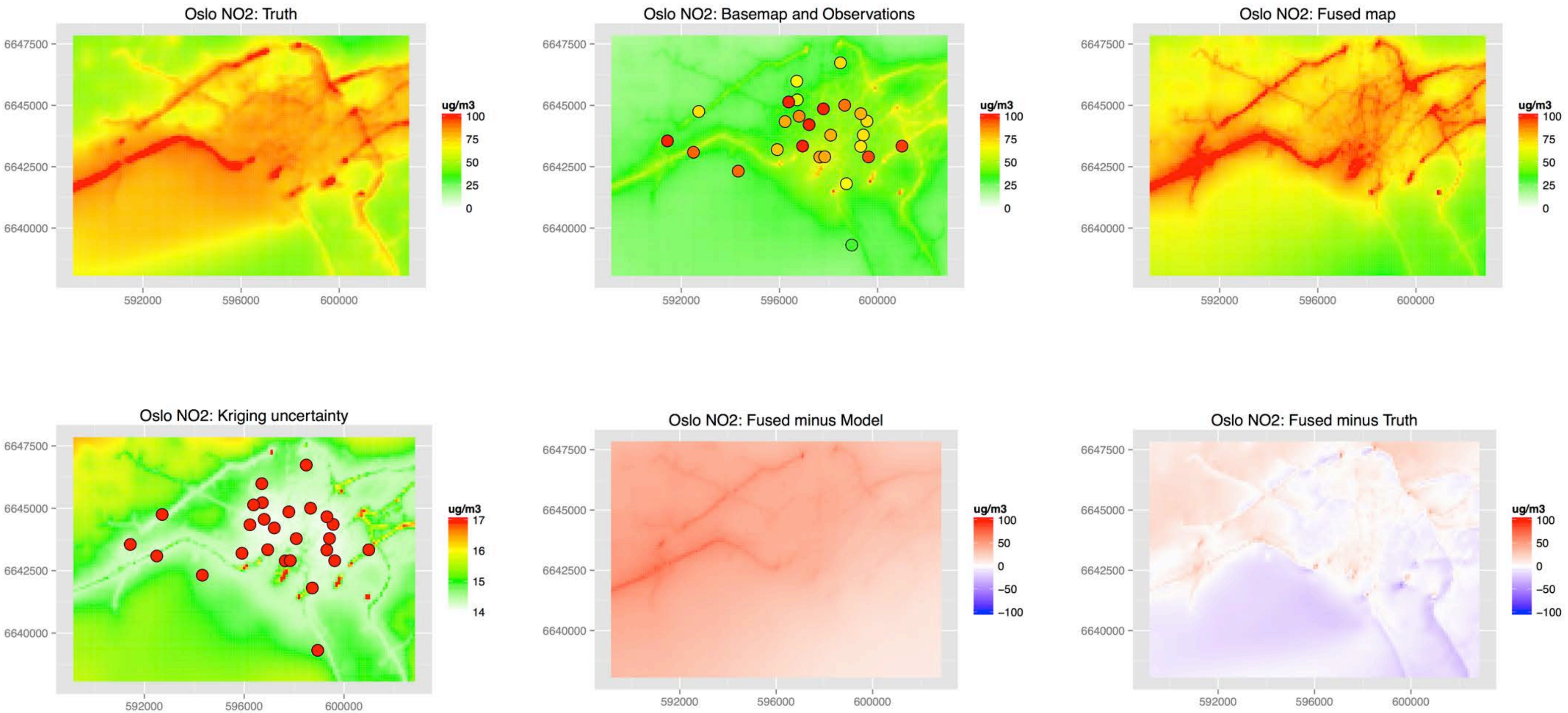




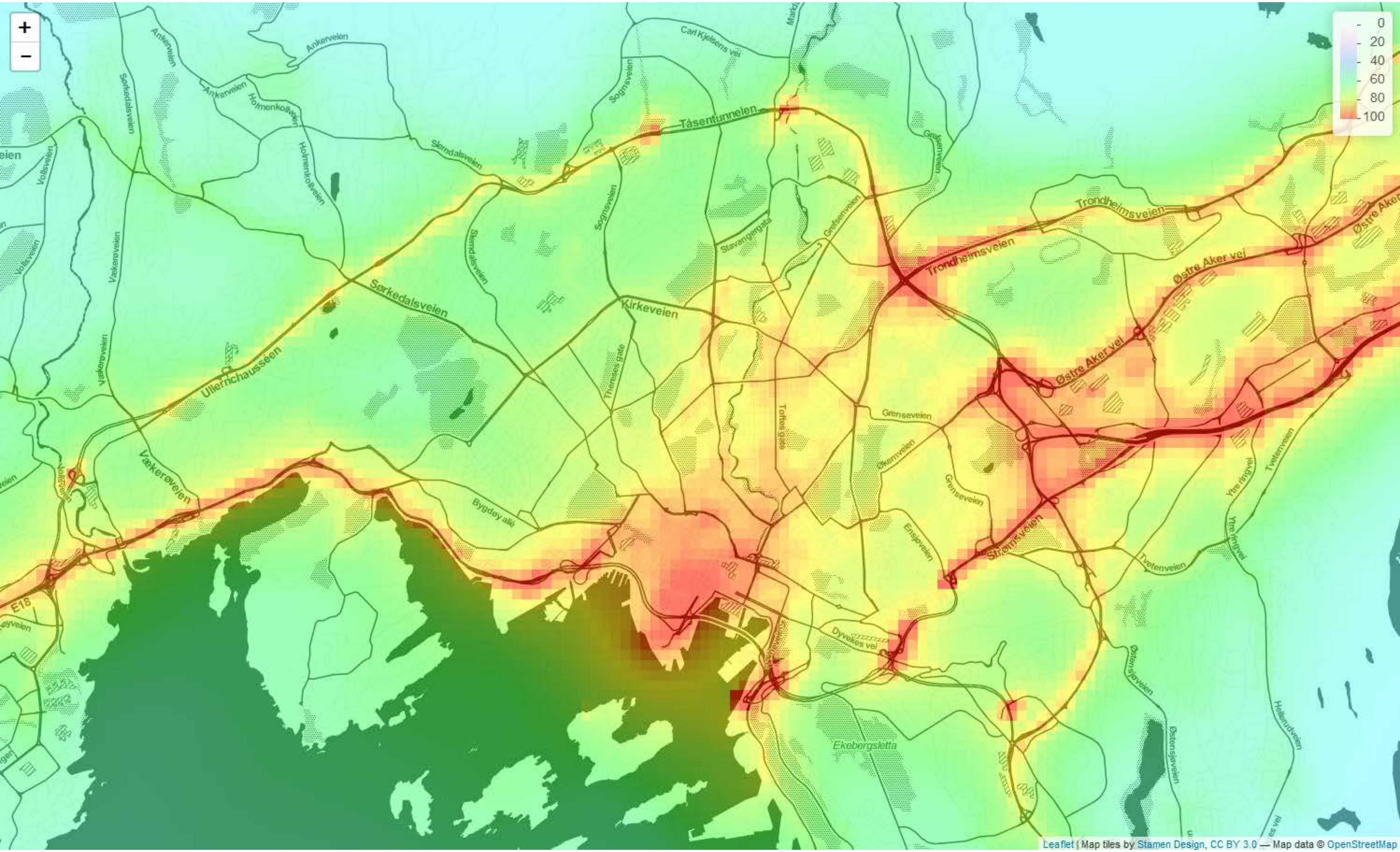


Schneider et al. (2015). Making sense of crowdsourced observations. EGU General Assembly 2015, Session ESSI1.10/SSS0.26, 15 April 2015, Vienna, Austria

A fusion example for Oslo



Schneider et al. (2015). Making sense of crowdsourced observations. ESA eo open science 2.0, 13 October 2015, Frascati, Italy

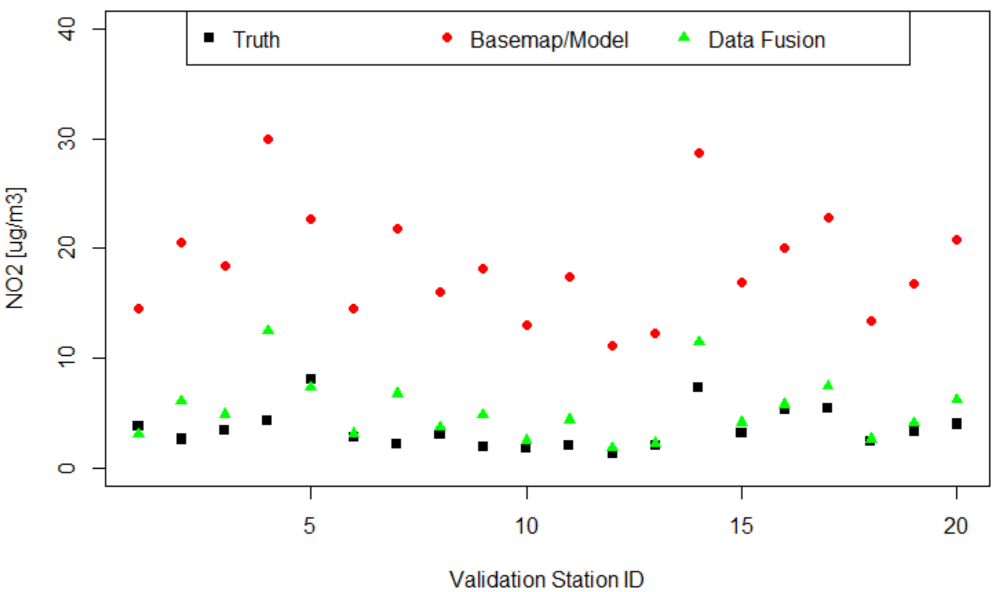


Typical example of a data fusion-based surface concentration field of NO_2 for Oslo, Norway, at 100 m spatial resolution.

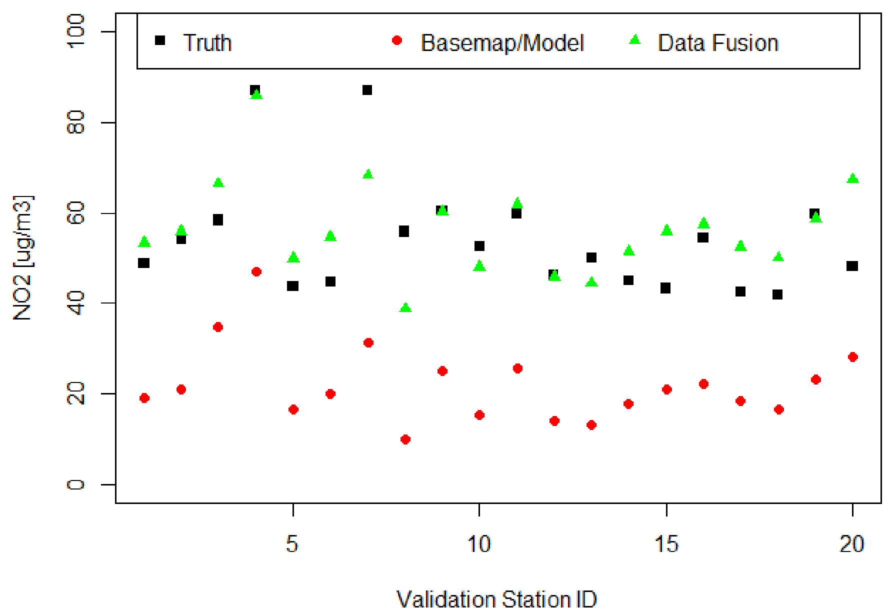
Validation against "Truth"

Validation sites are randomly selected throughout the image. Concentration values at these sites can be extracted from the truth, the basemap, and the fused result and compared.

Example 1: Basemap overestimates "truth"

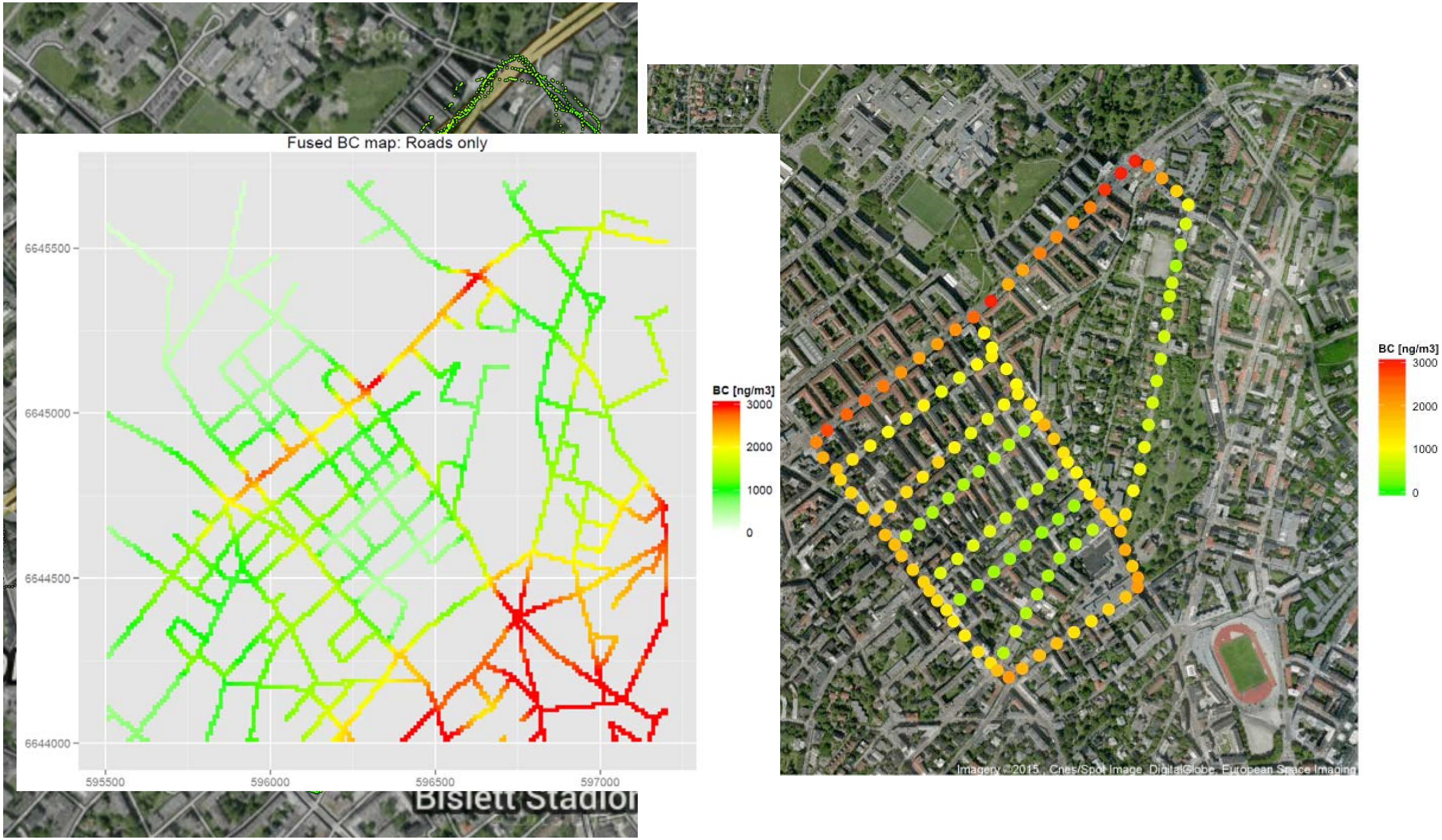


Example 2: Basemap underestimates "truth"



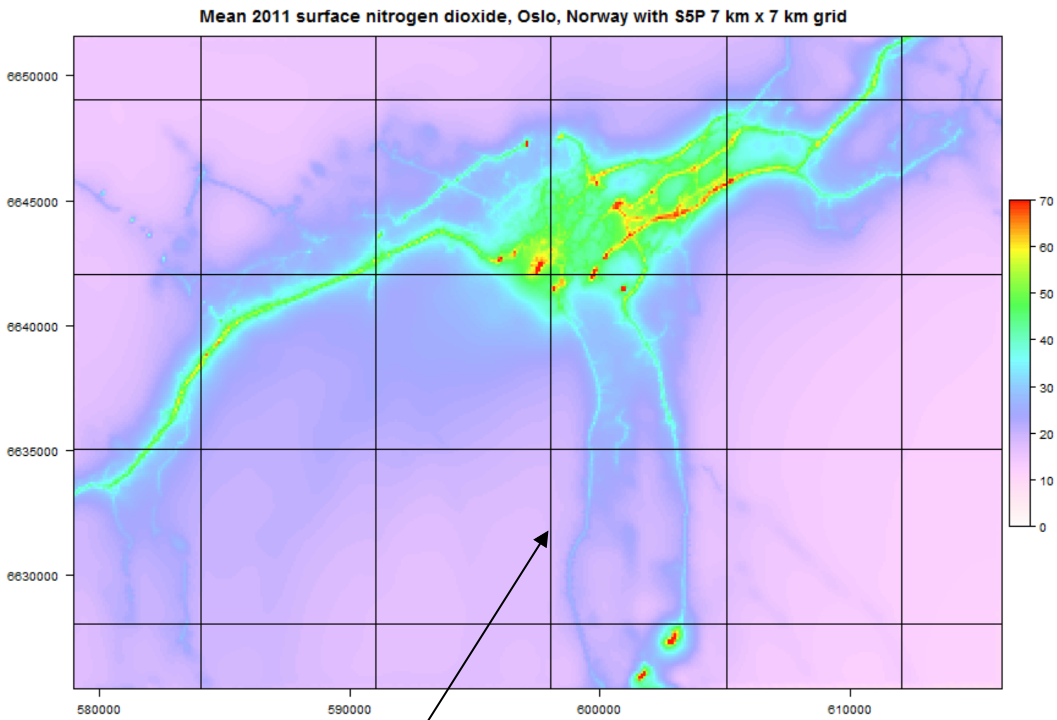
This shows that the method can predict the true concentration field quite well even in areas where no observations are available.

Data fusion of mobile measurements of Black Carbon



Schneider et al. (2015). Making sense of crowdsourced observations. ESA eo open science 2.0, 13 October 2015, Frascati, Italy

Applications for Sentinel-5P



7 km by 7 km grid of Sentinel-5P products for atmospheric composition

High-resolution urban air quality maps based on the combination of crowdsourced observations and model data provides sub-pixel information for Sentinel-5P

Could be used for:

- Validation/verification of S5P data (e.g. NO_2)
- Downscaling of the S5P products to higher spatial resolution using the fused map as proxy for spatial patterns

Conclusions

- We developed a technique for merging point-based crowdsourced observations of air quality with model information
  - **Geostatistics**-based: Builds upon decades of experience; best linear unbiased estimator; provides uncertainty estimates
  - **Fully automated** implementation: Can be run operationally in real time with large datasets
- Provides a much more **realistic estimate of true concentration field** than observations or model data alone
- Realistic high-resolution near real-time concentration fields in urban areas for the first time allow for **personalized air quality information**
  - "How much particulate matter will I breathe in if I ride my bike **from home to work right now**?"
  - "What **route** to work is the **least polluted/healthiest**?"
- In future, the resulting up-to-date concentration fields could be used for **validation and/or downscaling of Sentinel-5P** products
- Not just for air quality: Methodology is useful for most crowdsourcing applications where point observations need to be combined with model data (or other auxiliary information)
- A first step towards **making sense of highly distributed observations** in the age of *crowdsourcing, Citizen Science*, *ubiquitous sensing*, and *Big Data*

Thank you for your attention!

Contact: Philipp Schneider

Email: ps@nilu.no

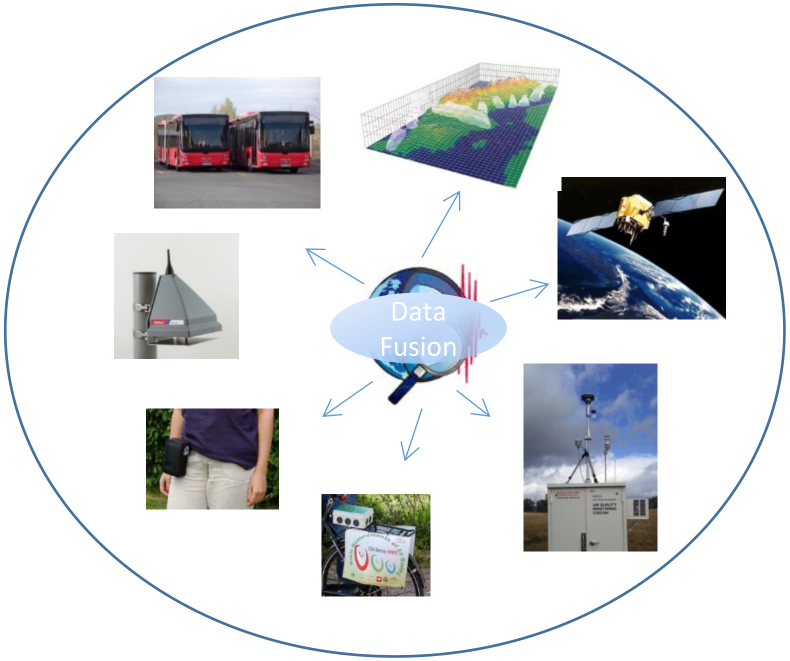
www.nilu.no

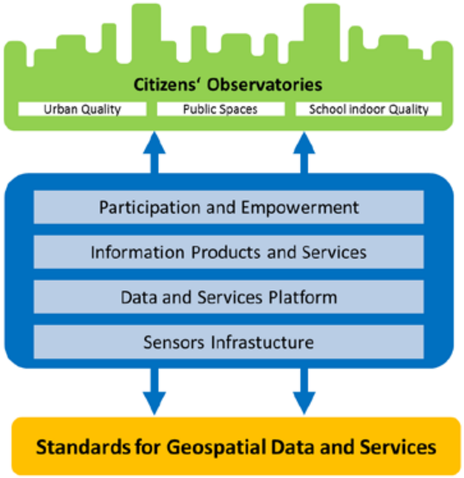
www.citi-sense.eu

Extra Slides

The CITI-SENSE project

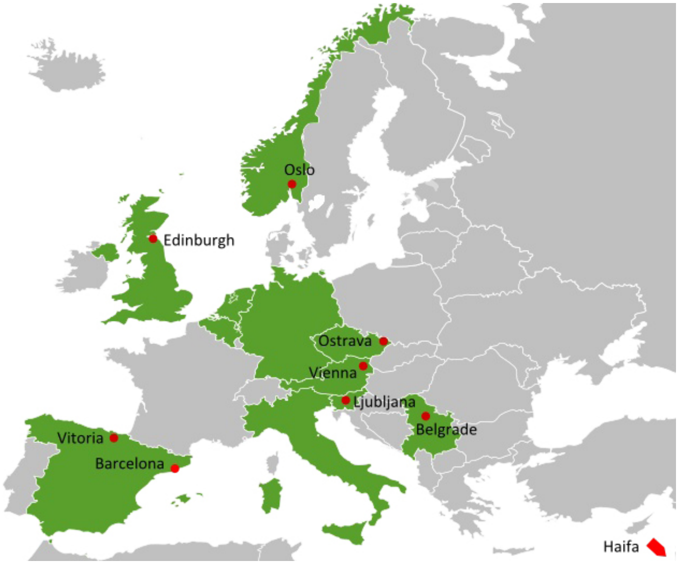
- Collaborative Project funded by FP7-ENV-2012
- 28 project partners from 12 countries (Europe, South Korea, and Australia)
- Objective: Development of sensor-based Citizen's Observatories for improving urban quality and for empowering citizens to
  - Contribute to and participate in environmental governance
  - Support and influence community and policy priorities and associated decision making
  - Contribute to the Global Earth Observation System of Systems (GEOSS)





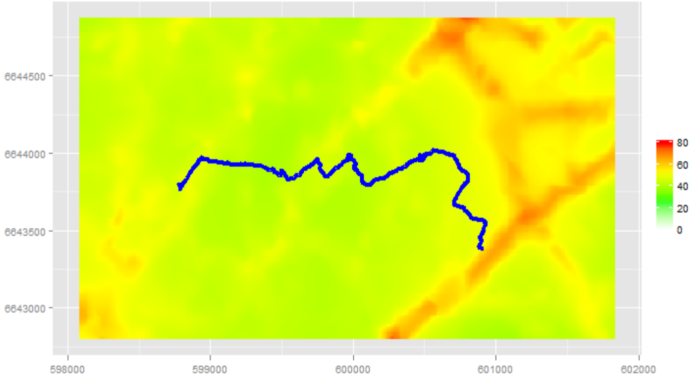
Observations

- There are many aspects to CITI-SENSE
  - Air quality observations using static and mobile sensor nodes
  - Indoor environment in schools
  - Public Spaces
- Here we focus on a network of static sensor nodes for air quality that are being deployed in various cities throughout Europe
  - Measuring the major air pollutants
  - Mounted at stakeholder's premises

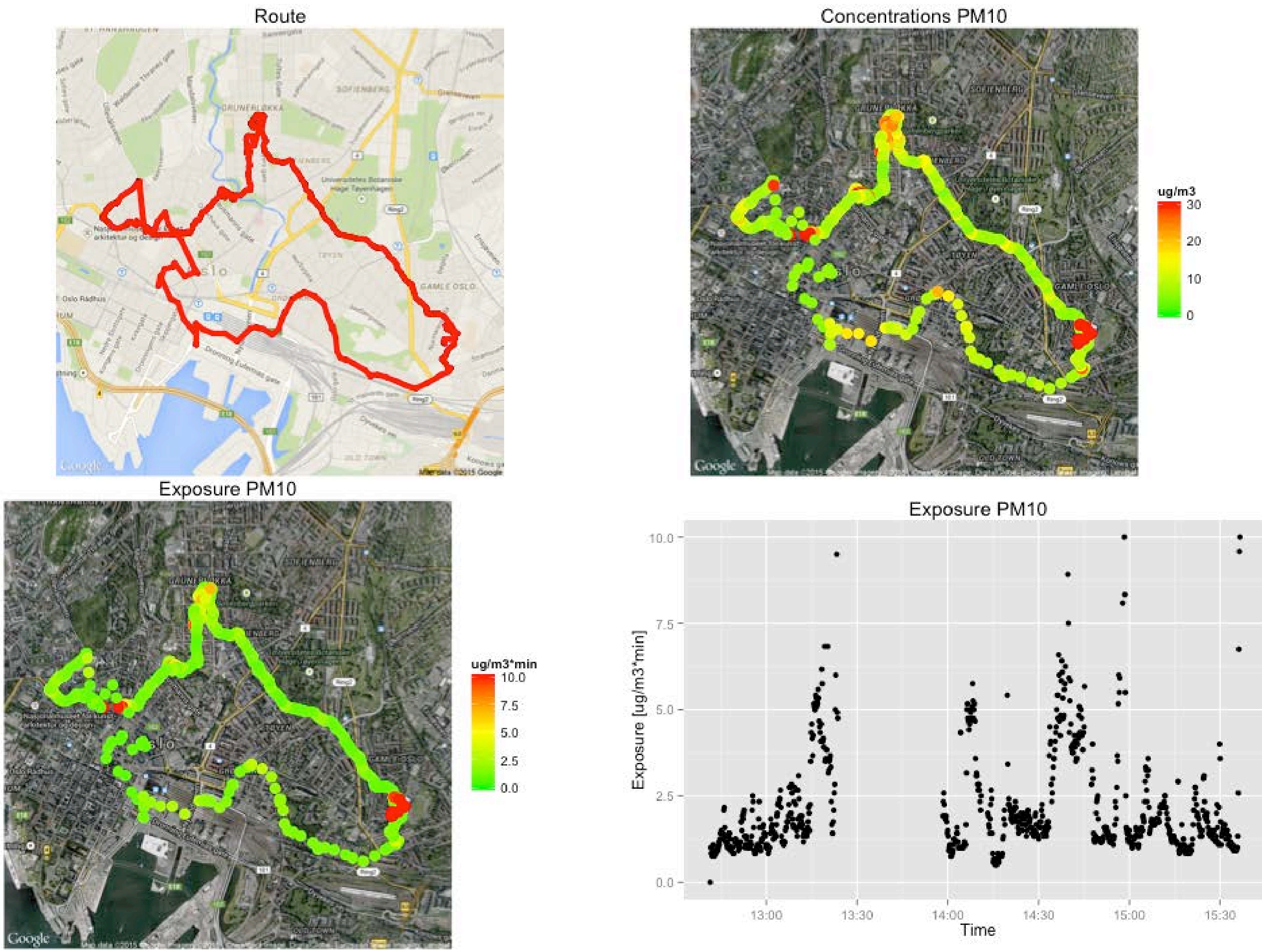


Towards personal exposure estimates

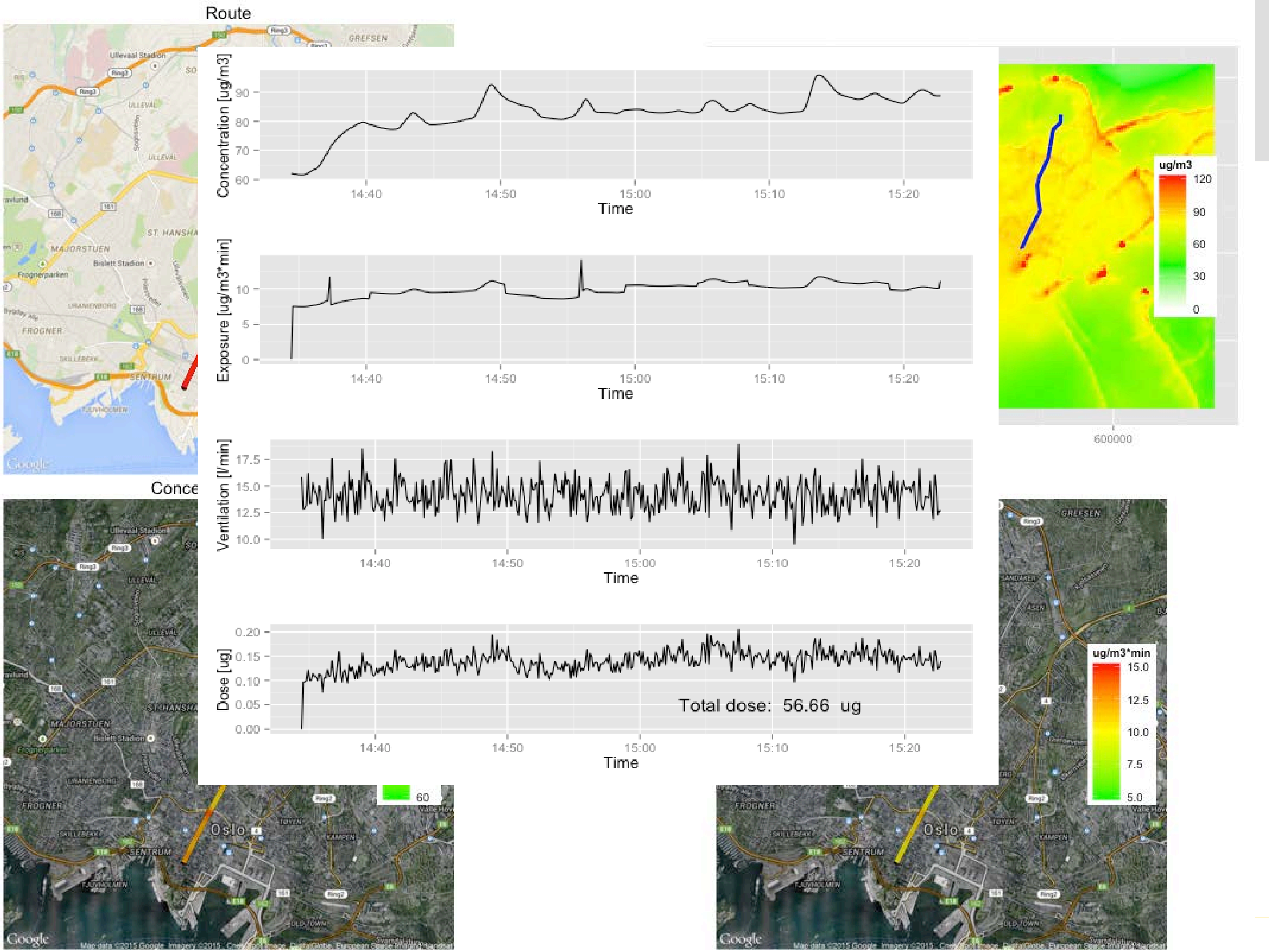
- There are two alternative ways for accomplishing personalized exposure/dose estimates
  - Approach 1: **Direct use of sensors**
    People move through the urban environment with portable sensors measuring concentrations
  - Approach 2: **Indirect use of sensors**
    Sensor data is used with model info and data fusion techniques to provide up-to-date air quality maps for the city -> these maps are then used to estimate exposure along a given track



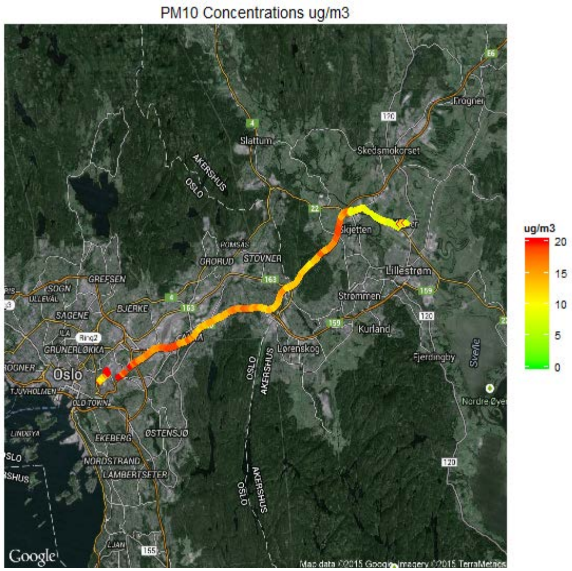
Sensor-based exposure and dose

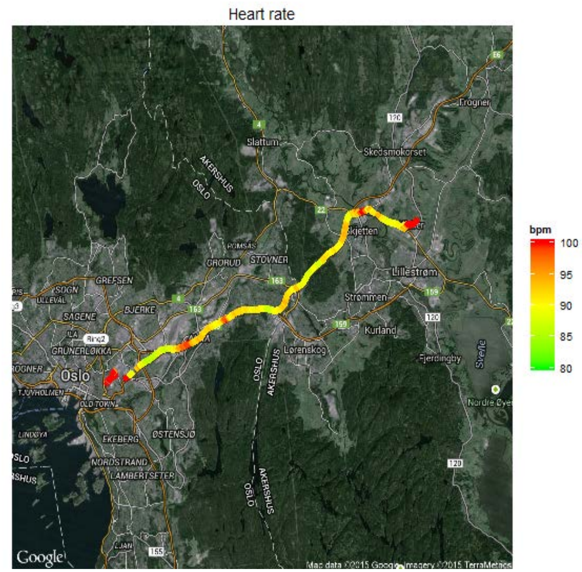


Schneider et al. (2015). Making sense of crowdsourced observations. ESA eo open science 2.0, 13 October 2015, Frascati, Italy


Route
Conce
Concentration [ug/m3]
Exposure [ug/m3*min]
Ventilation [l/min]
Dose [ug]
Time
14:40
14:50
15:00
15:10
15:20
Total dose: 56.66 ug
ug/m3
120
90
60
30
0
600000
ug/m3*min
15.0
12.5
10.0
7.5
5.0
Oslo
SENTRUM
FROGNER
MAJORSTUEN
ULLEVÅL
GREFSEN
Ullevaal Stadion
Bislett Stadion
Frognerparken
ST. HANSHA
URANIENBORG
SKILLEBEKK
TJUVHOLMEN
KAMPEN
TØYEN
OLD TOWN
Google

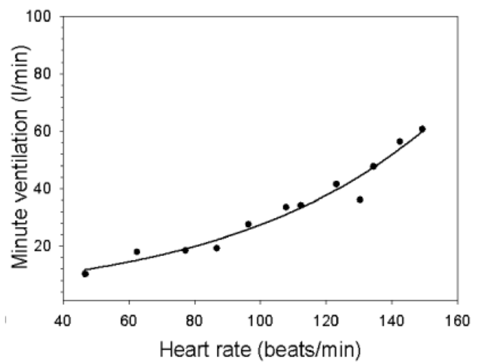
Heart rate -> Dose

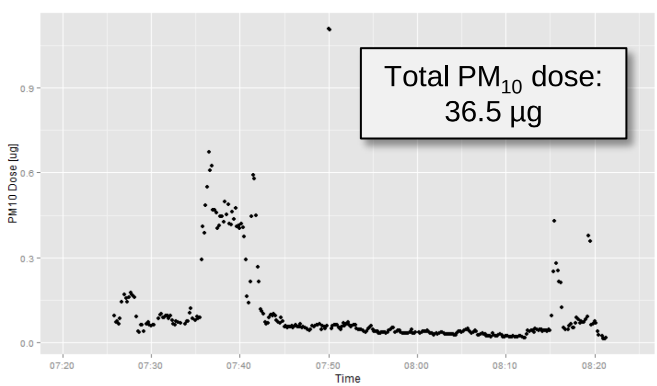




If we know the subject's hear rate we can compute the inhalation rate (ventilation) and the inhaled dose

Inhaled dose = Concentration × Ventilation × Duration





Total PM_{10} dose: 36.5 µg

Heart rate can either be

1. Measured by a heart rate monitor (mostly Approach 1)
2. Derived from accelerometer data (Approach 1+2)
3. Estimated by activity (mostly Approach 2)