European Network on New Sensing Technologies for Air Pollution Control and Environmental Sustainability - *EuNetAir*

COST Action TD1105

WGs and MC Meeting at ISTANBUL, 3-5 December 2014

Action Start date: 01/07/2012 - Action End date: 30/06/2016

Year 3: 1 July 2014 - 30 June 2015 (*Ongoing Action*)

Spatial Data Fusion of Observations from Low-Cost AQ Sensors and Models for Urban-Scale Air Quality Mapping

NILU

Philipp Schneider

WG3 Member and MC Member

Norwegian Institute for Air Research (NILU)
Kjeller, Norway

cost
EUROPEAN COOPERATION IN SCIENCE AND TECHNOLOGY

EuNetAir
European Network on New Sensing Technologies for Air-Pollution Control and Environmental Sustainability

COST is supported by the EU Framework Programme

European Science Foundation
ESF provides the COST Office through a European Commission contract

COST TD1105 WG Meeting
3-5 December 2014 – Istanbul, Turkey

Spatial Data Fusion of Observations from Low-Cost AQ Sensors and Models for Urban-Scale Air Quality Mapping

Philipp Schneider[1]
Nuria Castell[1]
William A. Lahoz[1]

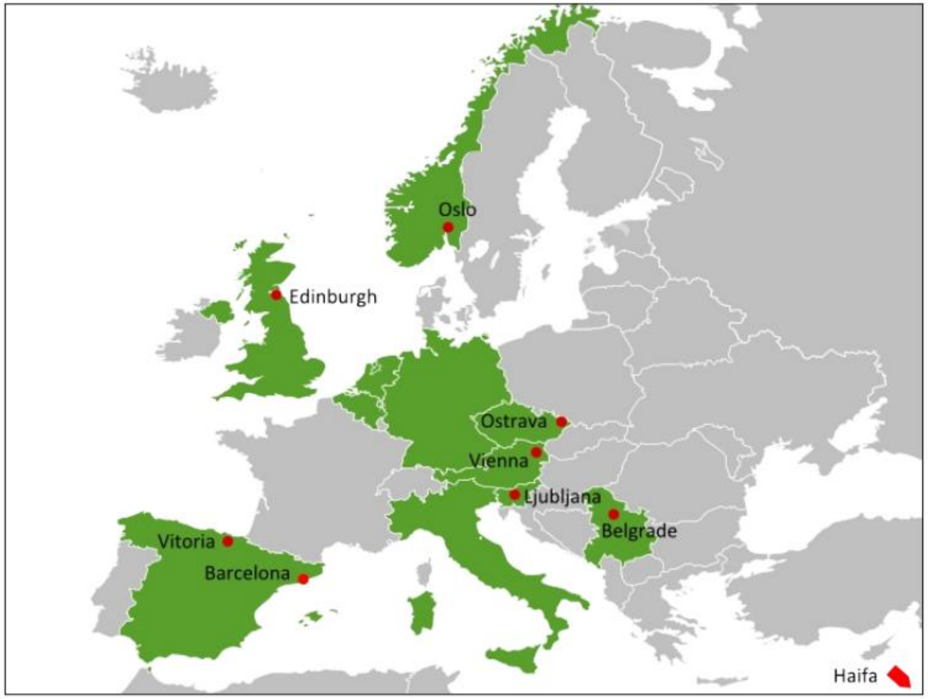
[1]NILU - Norwegian Institute for Air Research

Introduction

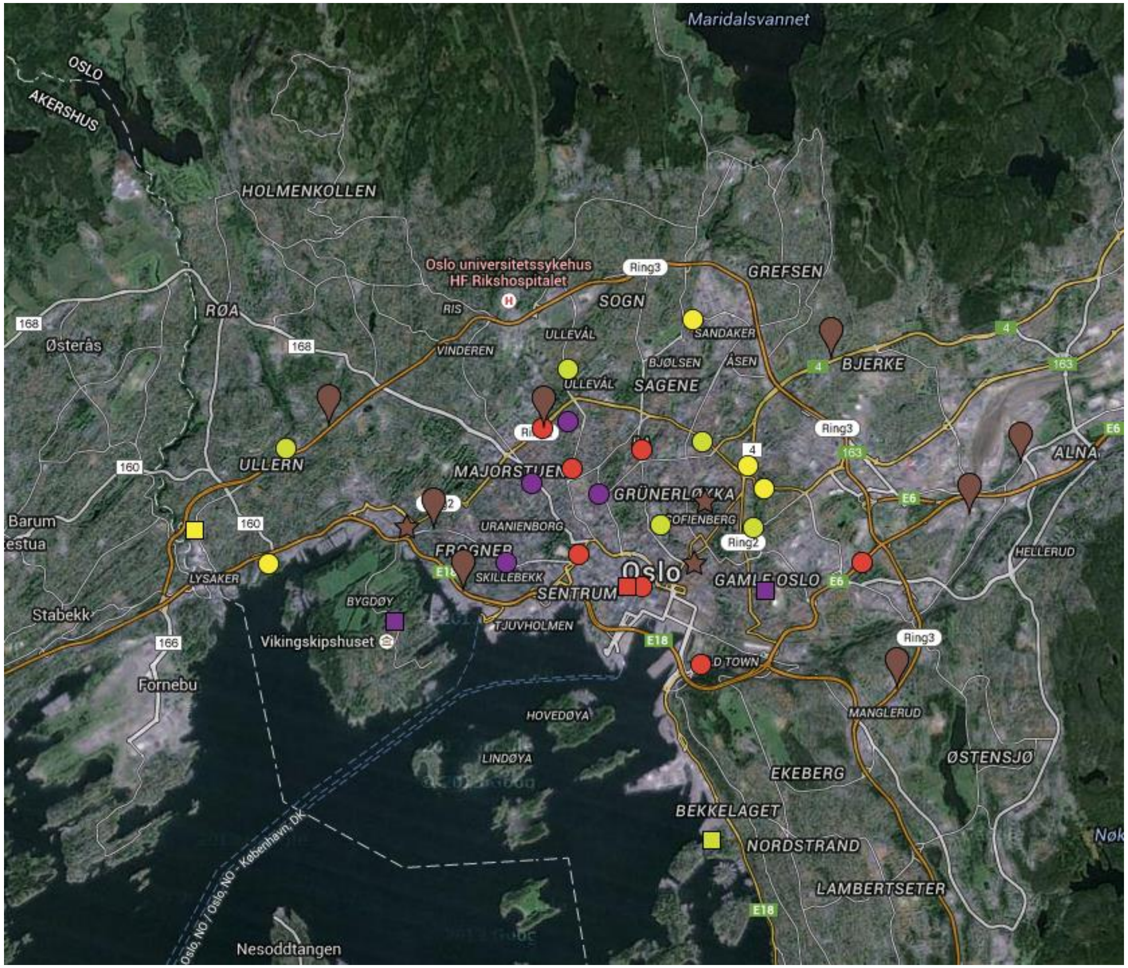
- Improving the mapping of urban scale air quality is one of the most promising potential applications of low cost microsensors
- Presently urban-scale mapping is very challenging due to low numbers of AQ reference stations
- Low cost AQ sensor can significantly increase the density of the monitoring network
- Even at higher deployment densities possible with low-cost sensors, realistic spatial mapping still requires city-scale model information to supply spatial patterns
- Combining model and sensor information can be achieved through data fusion (presented here) or data assimilation (in future)

The CITI-SENSE project

- Development of sensor-based Citizen's Observatories for improving the quality of life in cities
- Collaborative Project funded by FP7
- 27 project partners from Europe, South Korea, and Australia
- Case studies at 9 locations throughout Europe



CITI-SENSE: Oslo case study



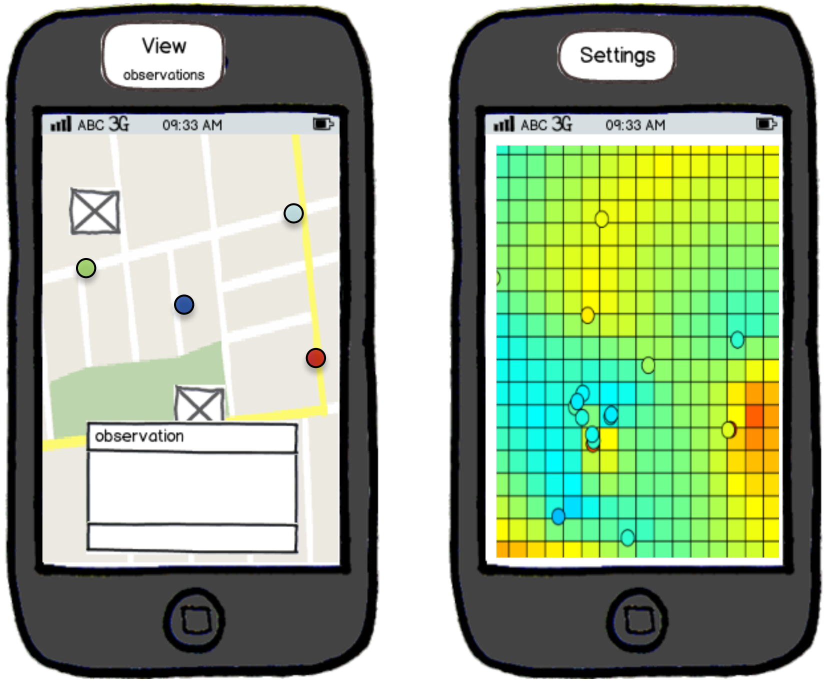
High-density network of static sensor nodes

21 Geotech sensor nodes will be installed throughout Oslo

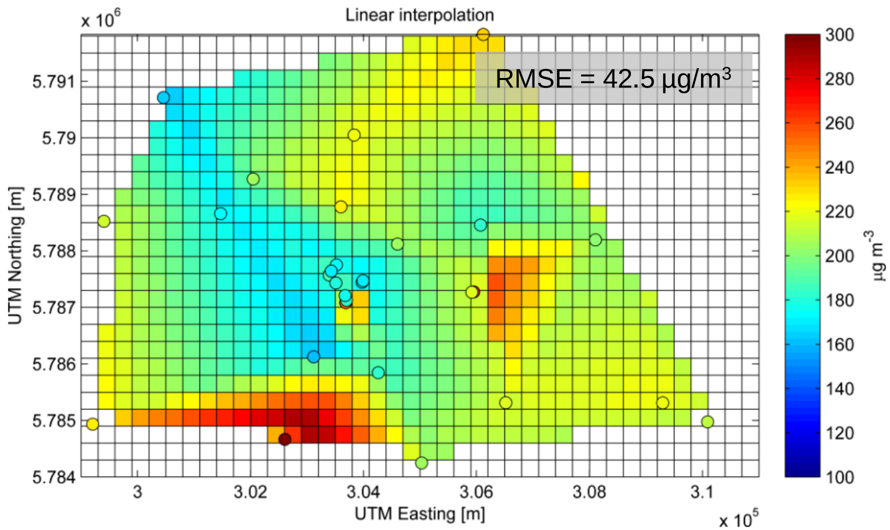
Schneider et al. 2014: Spatial Data Fusion of Observations from Low-Cost AQ Sensors and Models for Urban-Scale Air Quality Mapping. COST TD1105 WG Meeting, 3-5 Dec 2014, Istanbul, Turkey

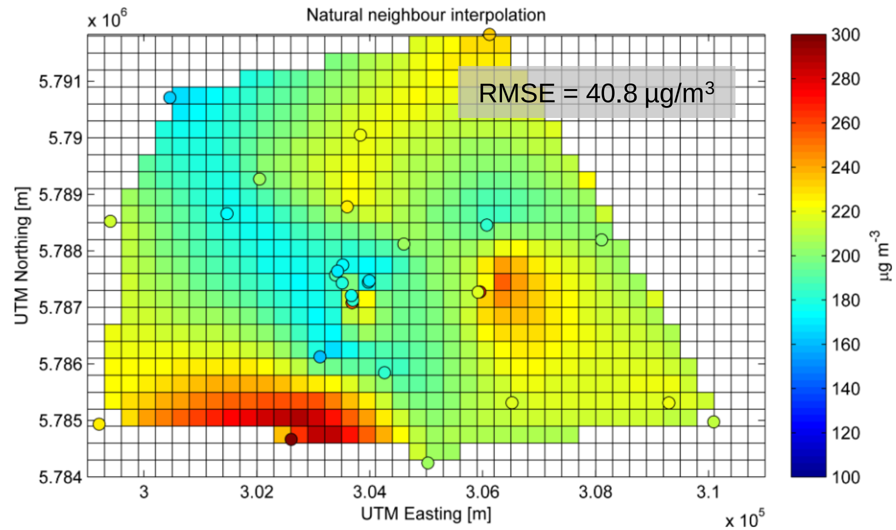
Motivation

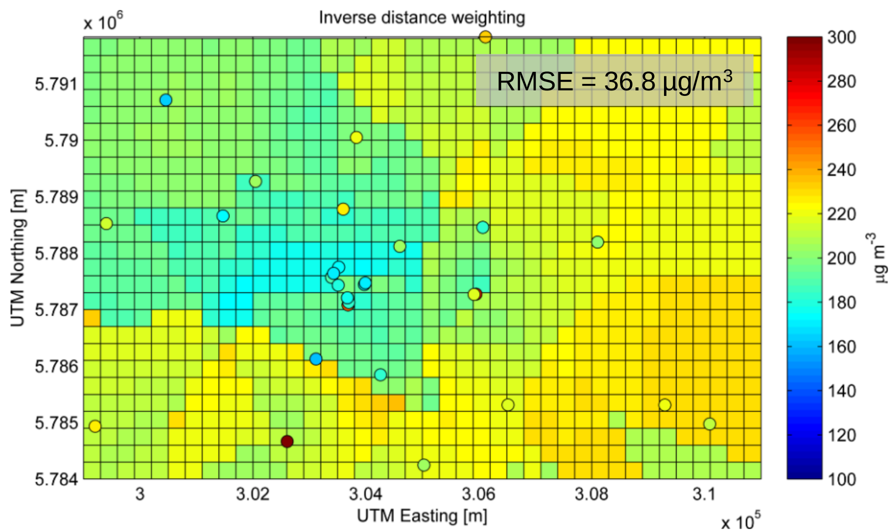
- CITI-SENSE requires a **continuous** spatial mapping technique that is **dynamic** and can be run **operationally** in **real-time** on a server **without human interaction**
- This is required for many applications that the users are interested in, such as
  - Planning the currently least polluted route through a city
  - Estimating personal exposure while moving through the city



Why not just interpolate?

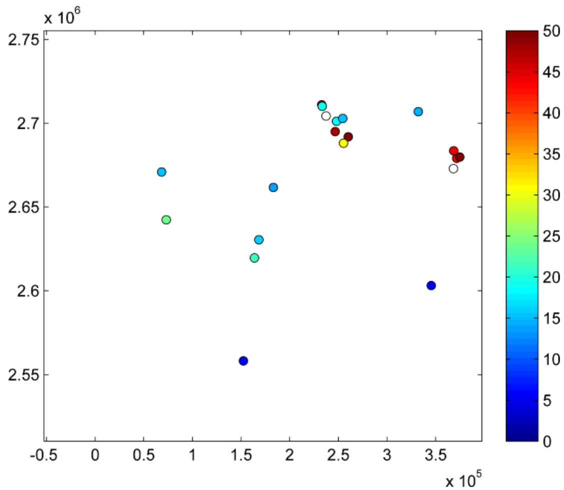






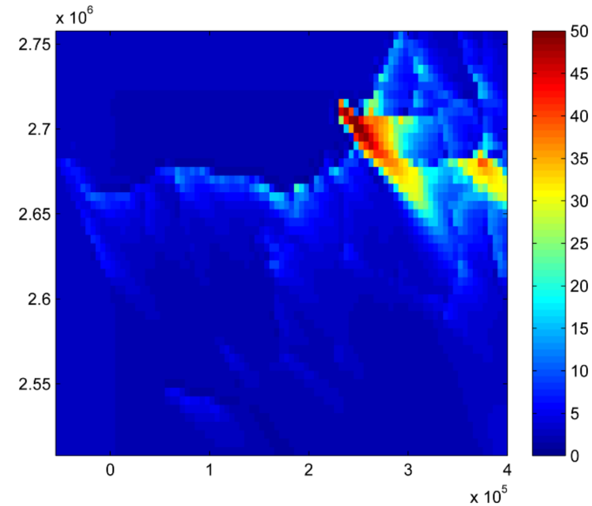
Schneider et al. 2014: Spatial Data Fusion of Observations from Low-Cost AQ Sensors and Models for Urban-Scale Air Quality Mapping. COST TD1105 WG Meeting, 3-5 Dec 2014, Istanbul, Turkey

Data fusion: Basic Premise

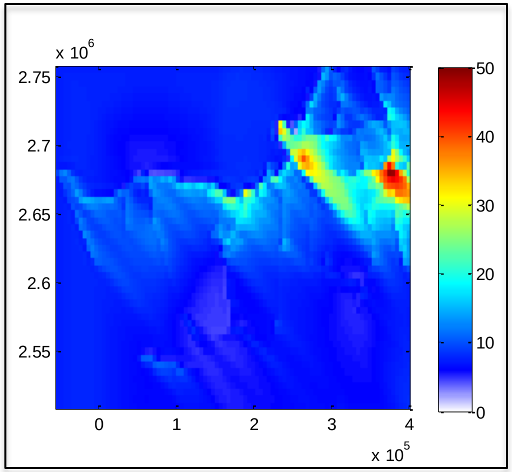


Observations

DATA FUSION



Modelling results or other auxiliary data



Combined map

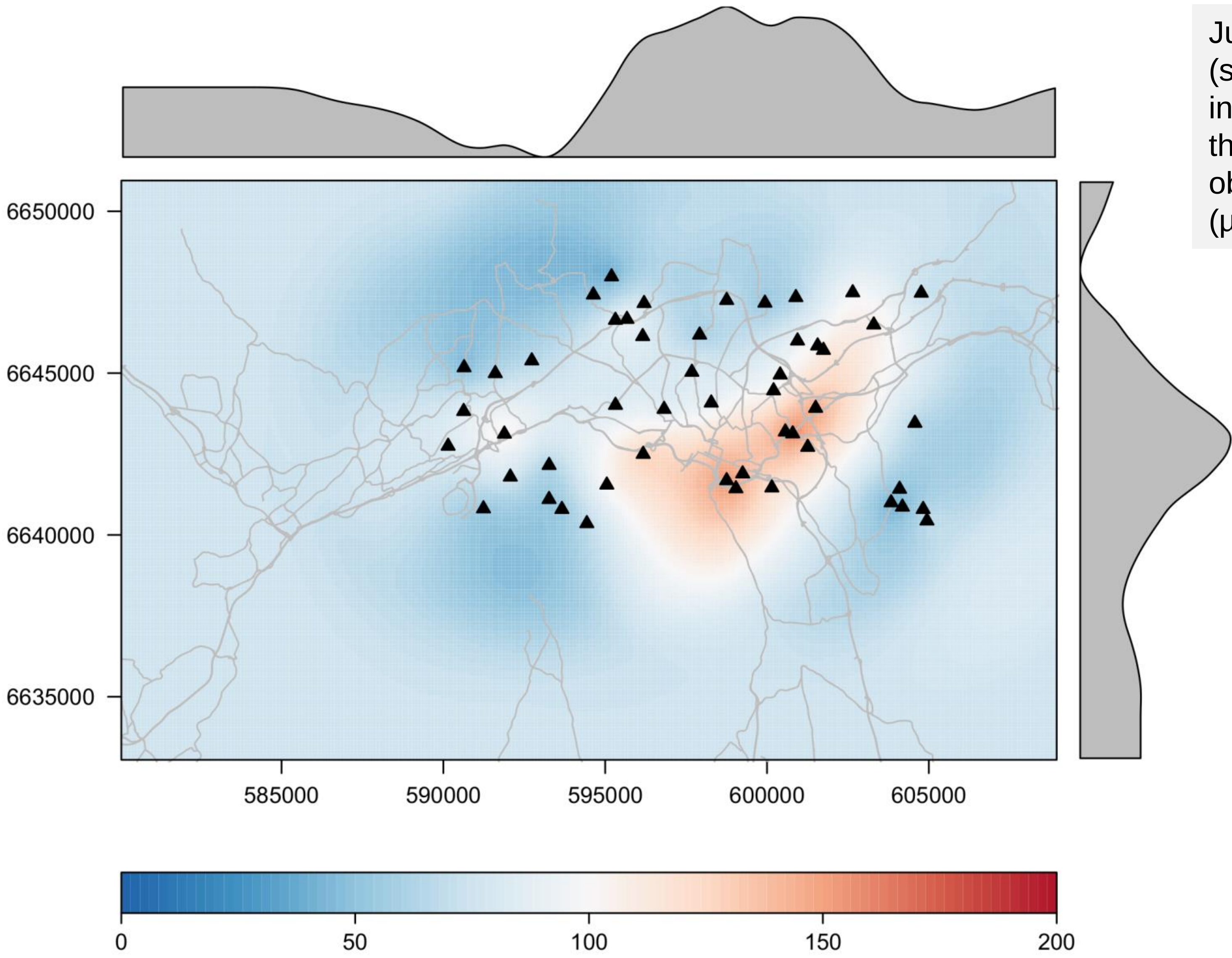
Data fusion (as a subset of data assimilation) creates a value-added product by

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

Data fusion method used here provides a combined concentration field by separately interpolating the observational residuals from a regression model and then combining both.

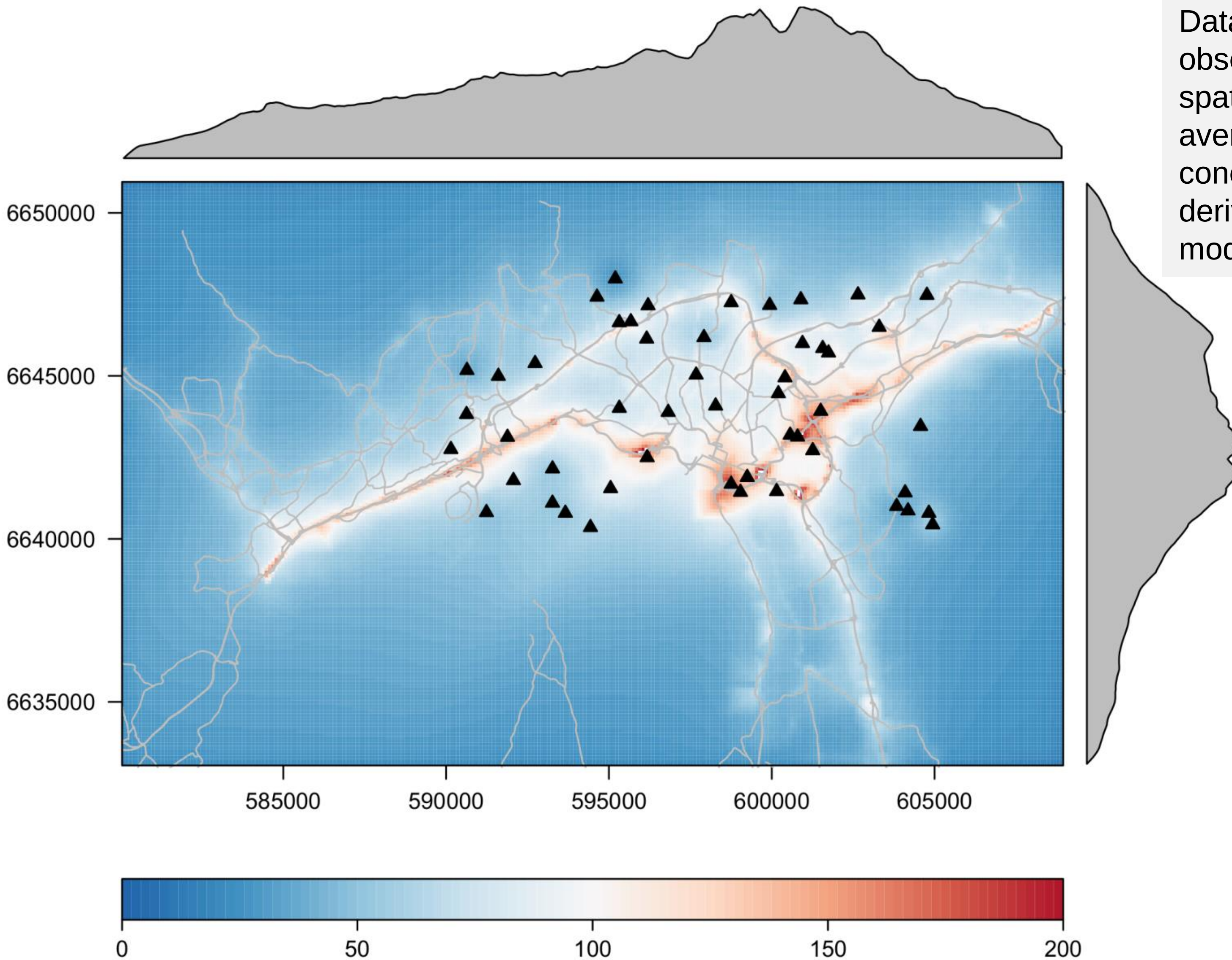
Schneider et al. 2014: Spatial Data Fusion of Observations from Low-Cost AQ Sensors and Models for Urban-Scale Air Quality Mapping. COST TD1105 WG Meeting, 3-5 Dec 2014, Istanbul, Turkey

Interpolation of just observations



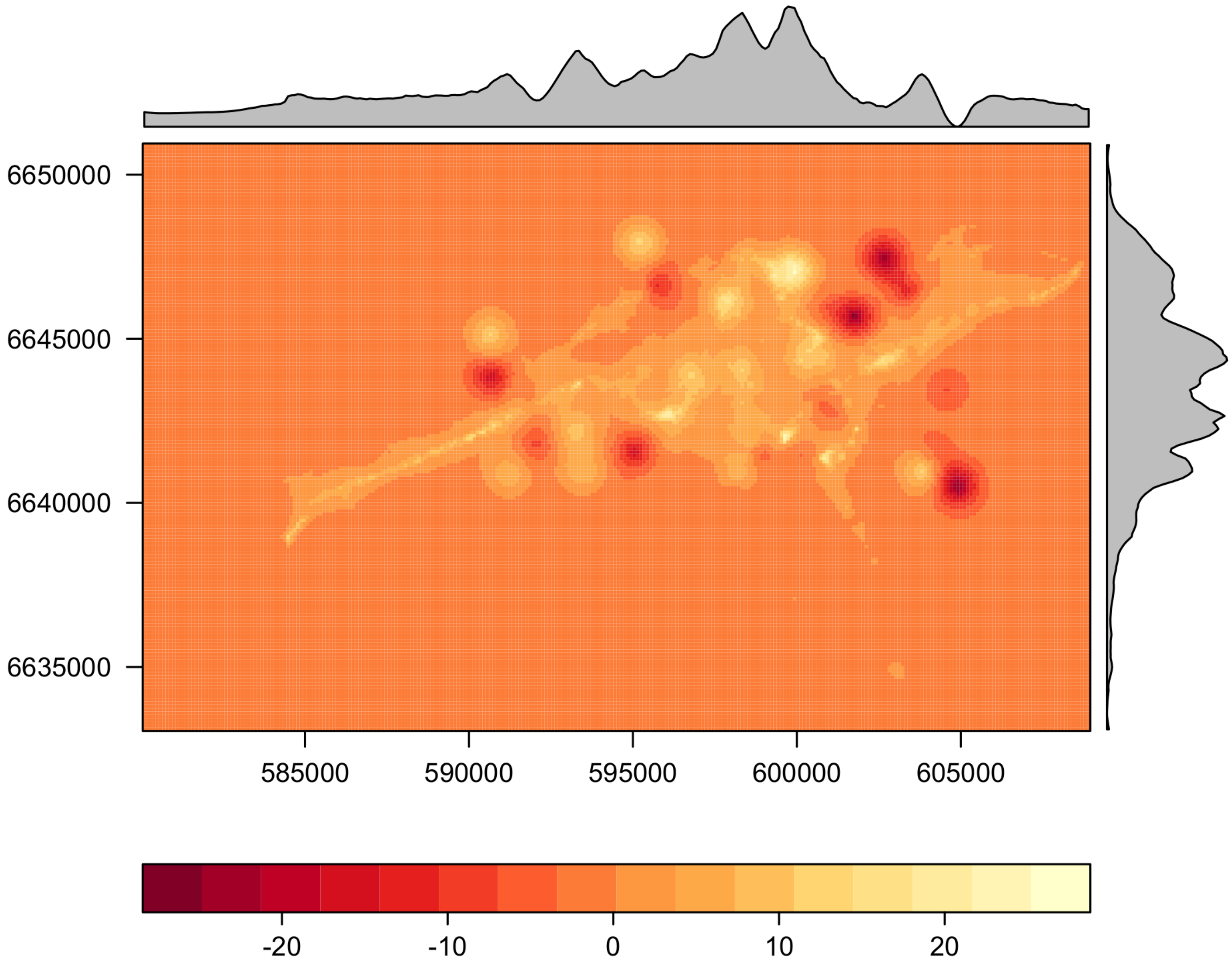
Just kriging (spatial interpolation) of the NO_2 observations (µg/m³)

Data fusion



Data fusion of the observations with a spatial proxy (annual average NO_2 concentration ($\mu g/m^3$) derived from EPISODE model ouput)

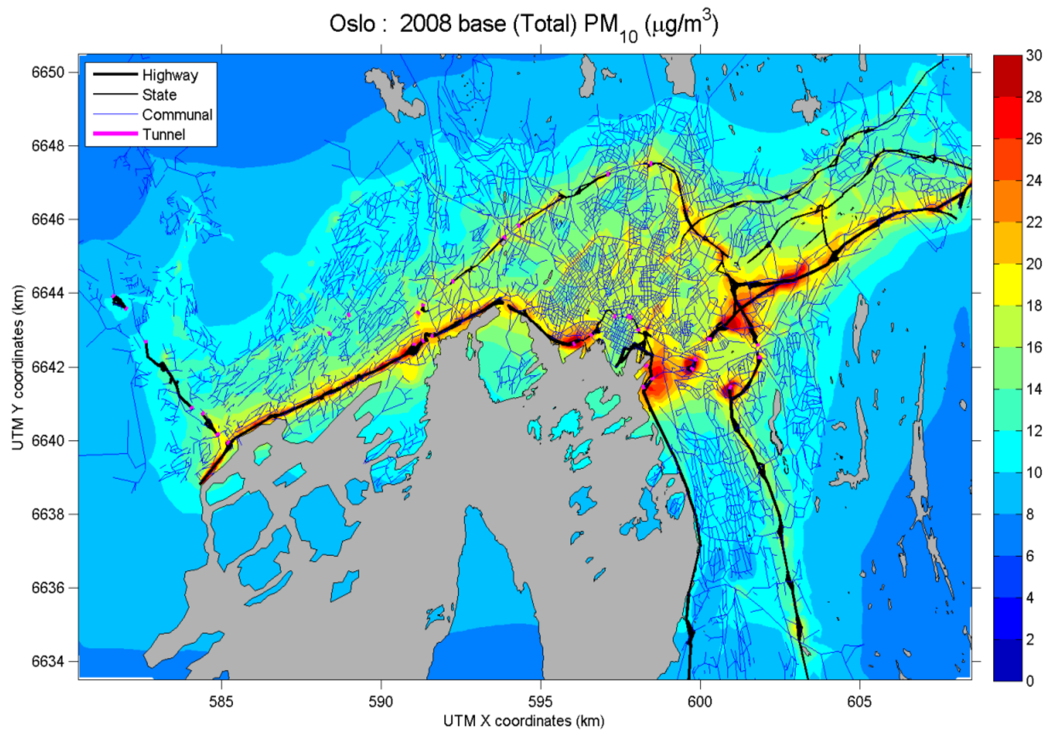
Data fusion



Difference between original proxy concentration grid for NO2 and the concentration grid coming out of the data fusion.

Spatial auxiliary (proxy) data

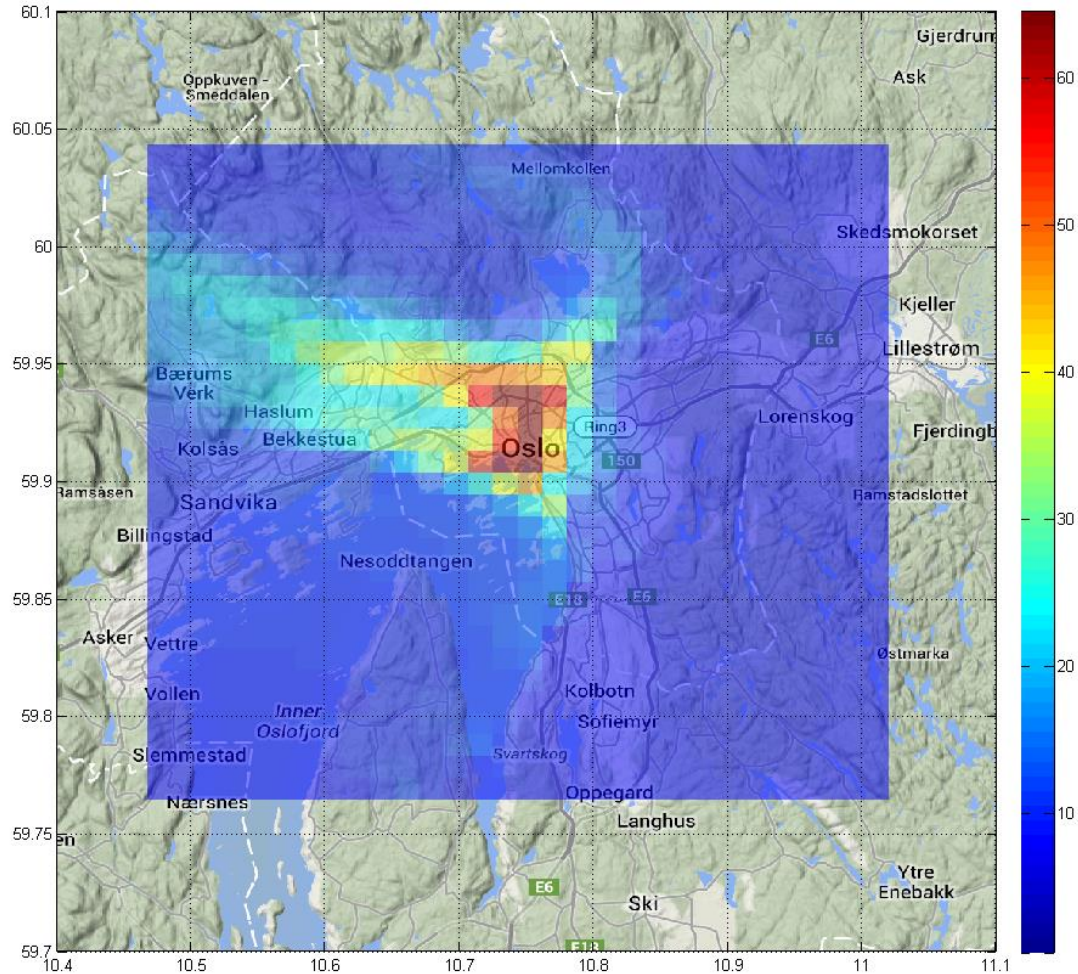
- Can be any spatially exhaustive dataset that is related to the observation
- For example a concentration map created through LUR modelling
- Can also be output from a high-resolution dispersion model
- Or all of the above…



High-resolution map of PM_{10} in Oslo from the EPISODE dispersion model. These kind of maps are ideally suited as a spatially distributed auxiliary dataset.

The EPISODE model

- Developed by Slørdal et al. (2008)
- Three-dimensional, combined Eulerian/Lagrangian air pollution dispersion model, developed at NILU
- Main focus on urban and local-to-regional scale applications
- Provides gridded fields of ground-level hourly average concentrations
- Spatial resolution down to 100m (but usually run a 1 km)
- Time step between 10 s and 300 s
- Schemes for advection, turbulence, deposition, and chemistry



Example output for NO2 from the EPISODE model over Oslo, here at 1 km spatial resolution.

The EPISODE model: Input Data

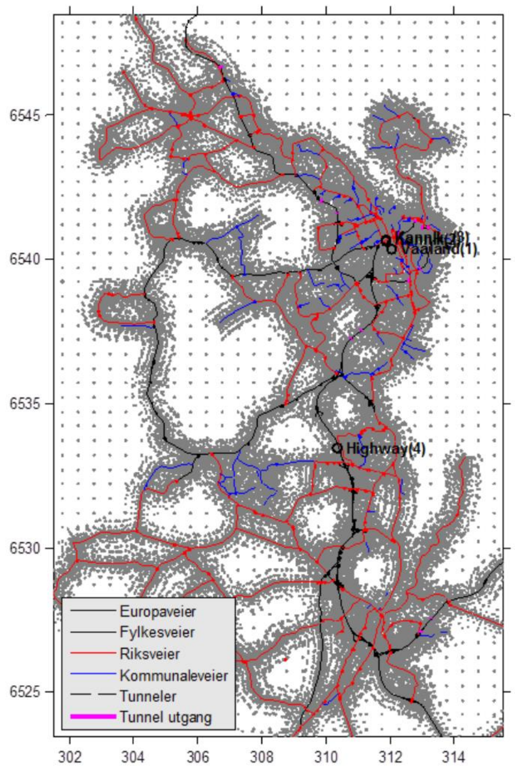
- **Emissions**
  Usually given as hourly values for each grid cell (area sources), road link (line sources), or stack (point sources)
- **Meteorology**
  Can be either given as gridded fields, for example from the output of existing meteorological models, or as the observations from a meteorological station within the domain
- **Topography**
  Given as a gridded field with the elevation for each grid cell. Important when only a meteorological station is used rather than gridded meteorological fields
- **Initial conditions**
  Initial concentrations (3D grid) for each pollutant must be provided by the user (usually set equal to zero). Final concentrations from a previous run may be used as initial concentrations for a new run.
- **Boundary conditions**
  Hourly background (boundary) concentrations for each pollutant must be provided by the user.

Preparing input data using EPISODE

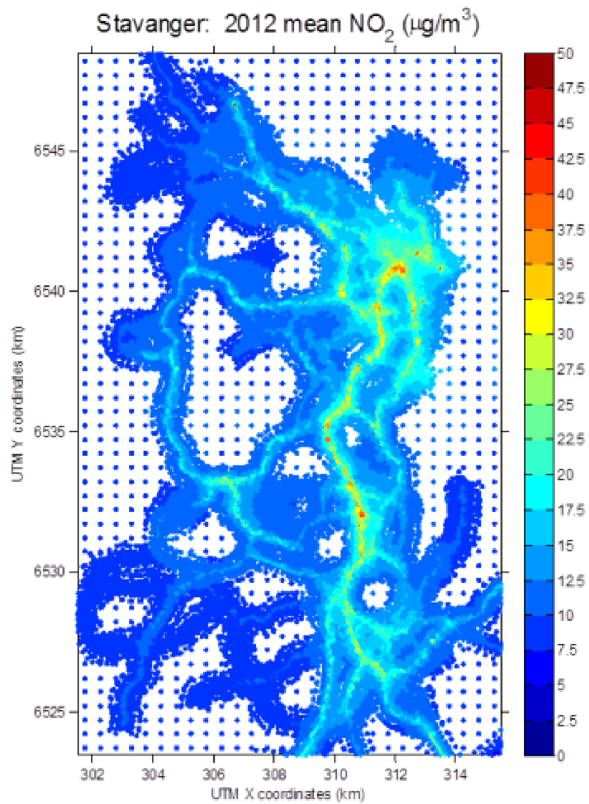
- Currently EPISODE is run at 1 km spatial resolution
- This limitation is mostly due to the availability of gridded emissions
- Spatial resolution required for CITI-Sense applications (e.g. personal exposure along track): on the order of 100 m
- We use a multi-step downscaling procedure to obtain gridded concentration fields with 100 m spatial resolution from EPISODE (developed by Denby et al. 2013)

Preparing input data using EPISODE

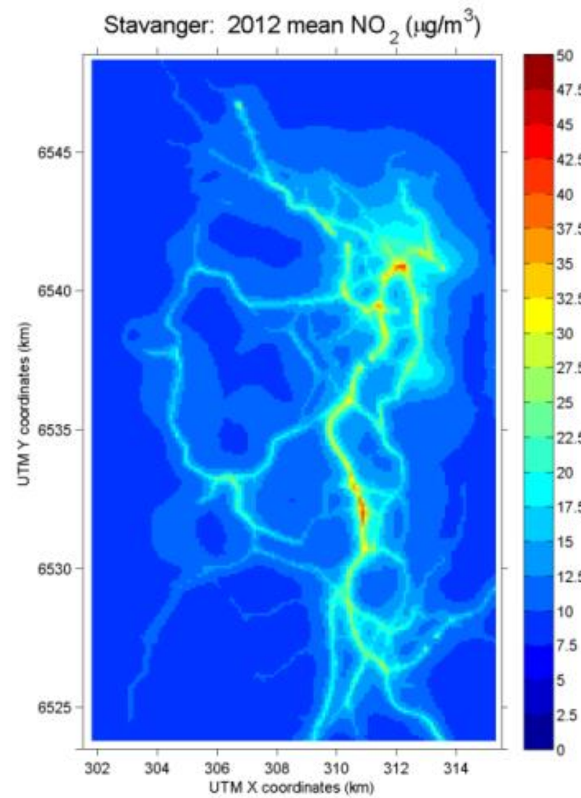
Output of model results at ~20000 receptor points distributed along roads. These concentrations are then subsequently mapped out on a regular grid.



Receptor points locations

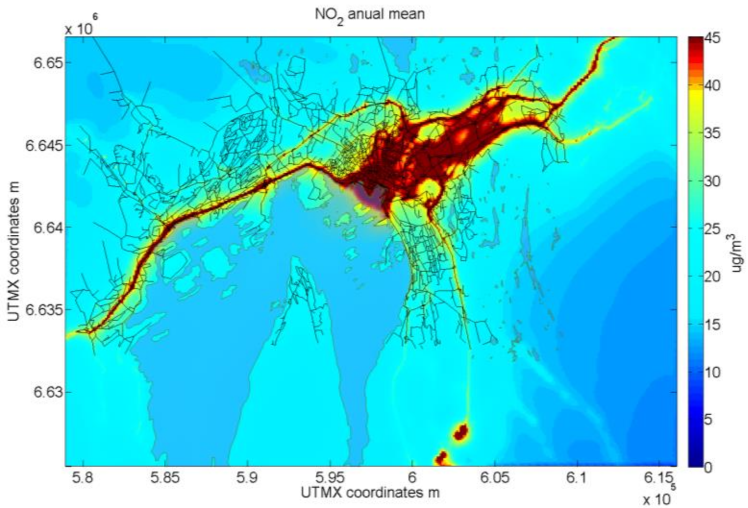


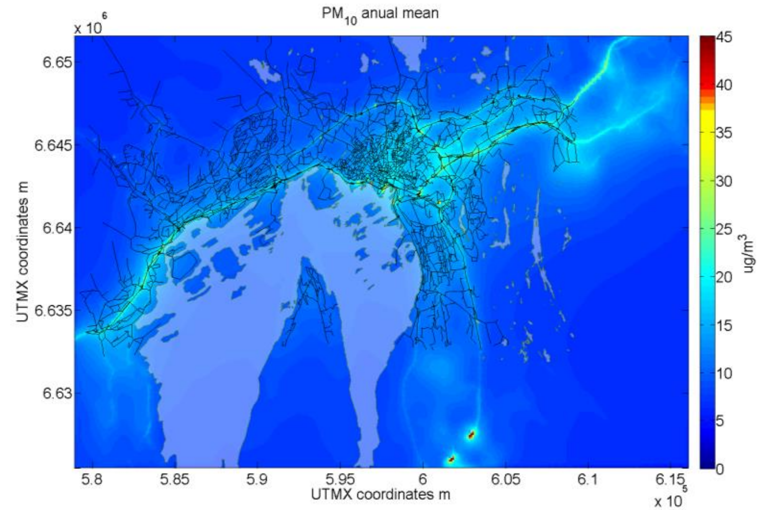
Modeled concentrations at receptor points



Final interpolated map of annual mean

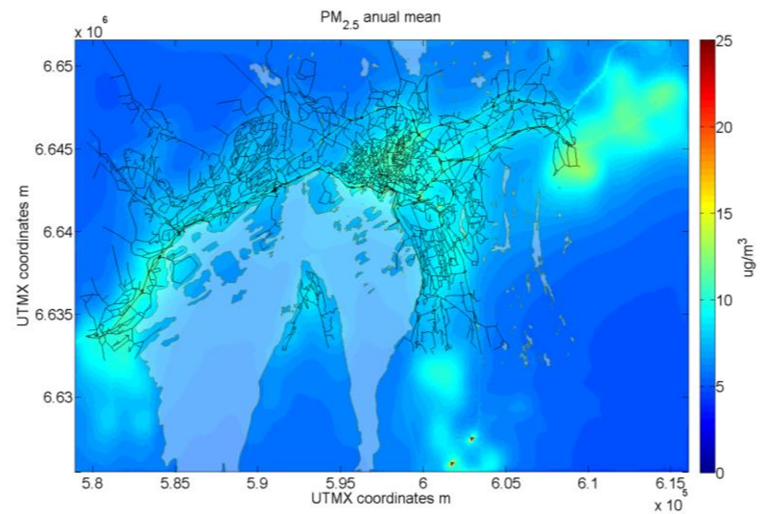
Preparing input data using EPISODE



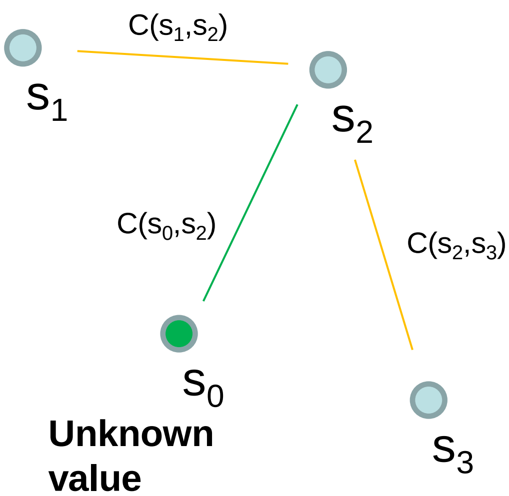


EPISODE dispersion modelling results. Annual mean concentrations for NO2, PM10, and PM2.5

Used as spatial proxies for near real-time data fusion within CITI-Sense



Kriging: Basic theory



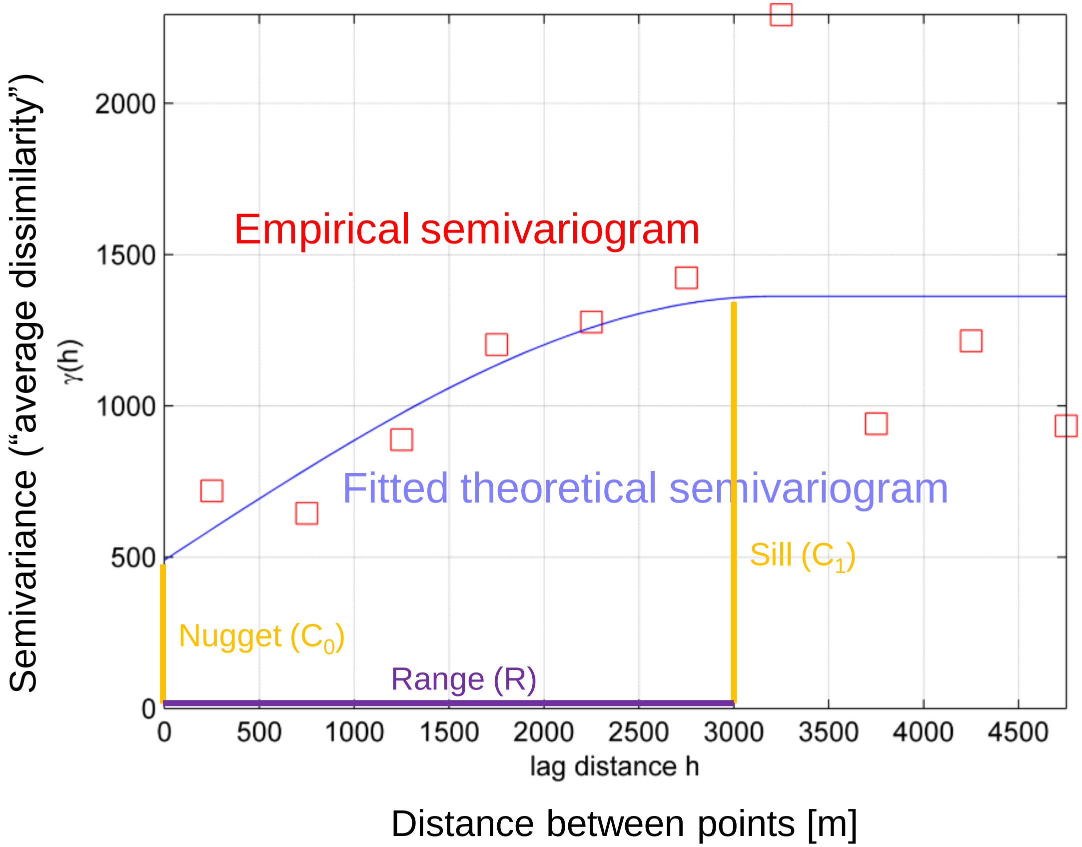
$$\hat{Z}(s_0) = \sum_{i=1}^{N} \boxed{w_i} Z(s_i)$$

$Z(s_i)$ = measured value at i-th location
w_i = unknown weight for observation at i-th location
s_0 = prediction location
N = number of observations

$$\begin{bmatrix} C(\mathbf{s}_1,\mathbf{s}_1) & \cdots & C(\mathbf{s}_1,\mathbf{s}_n) & 1 \\ \vdots & & \vdots & \vdots \\ C(\mathbf{s}_n,\mathbf{s}_1) & \cdots & C(\mathbf{s}_n,\mathbf{s}_n) & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}^{-1} \cdot \begin{bmatrix} C(\mathbf{s}_0,\mathbf{s}_1) \\ \vdots \\ C(\mathbf{s}_0,\mathbf{s}_n) \\ 1 \end{bmatrix} = \begin{bmatrix} w_1(\mathbf{s}_0) \\ \vdots \\ w_n(\mathbf{s}_0) \\ \varphi \end{bmatrix}$$

The Ordinary Kriging (OK) system: Used for calculating the weights w_i

Kriging: Covariance modeling



Semivariance γ(h)
Measure of average dissimilarity between observations as a function of their separation in distance and direction. Semivariance and covariance are closely linked.

$$\gamma(\mathbf{h}) = \frac{1}{2} E\left[\left(z(\mathbf{s}_i) - z(\mathbf{s}_i + \mathbf{h})\right)^2\right]$$

$$C(\mathbf{h}) = C_0 + C_1 - \gamma(\mathbf{h})$$

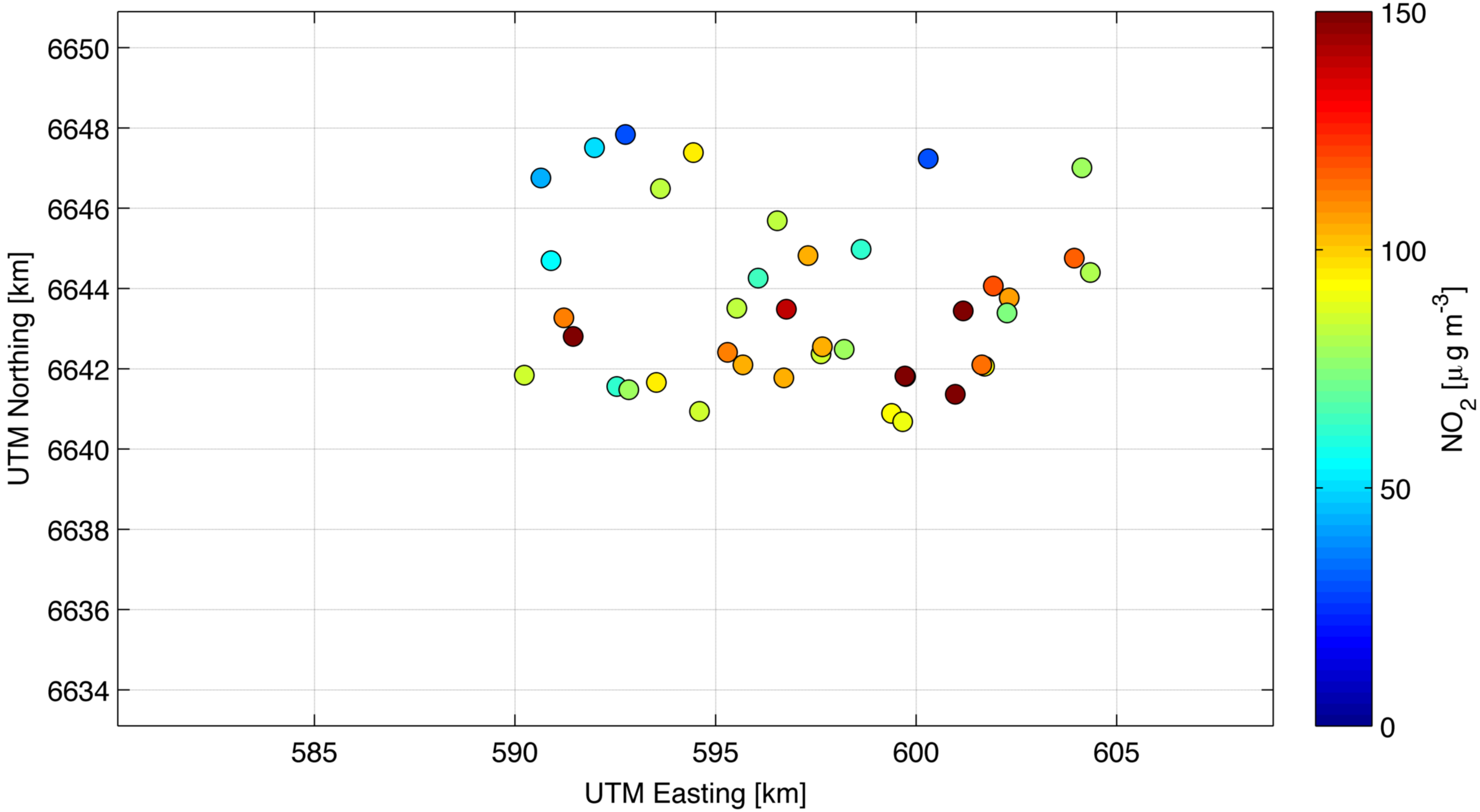
Many theoretical semivariogram models exist. Most used are

- Spherical
- Gaussian
- Exponential

$$\gamma(\mathbf{h}) = \begin{cases} 0 & \text{if} \quad |\mathbf{h}| = 0 \\ C_0 + C_1 \cdot \left[1 - e^{-\left(\frac{\mathbf{h}}{R}\right)}\right] & \text{if} \quad |\mathbf{h}| > 0 \end{cases}$$

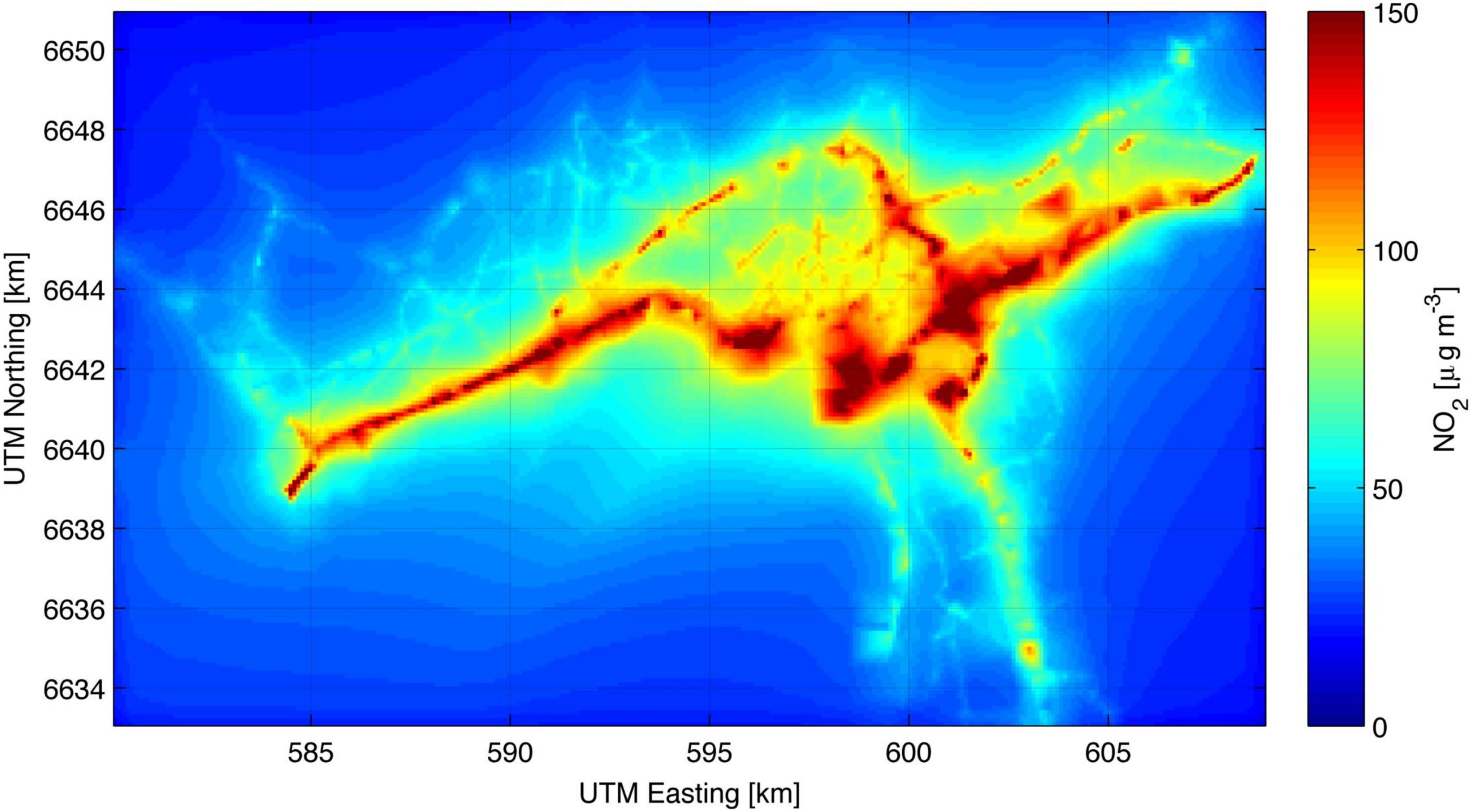
Schneider et al. 2014: Spatial Data Fusion of Observations from Low-Cost AQ Sensors and Models for Urban-Scale Air Quality Mapping. COST TD1105 WG Meeting, 3-5 Dec 2014, Istanbul, Turkey

Oslo Data Fusion Example: Observations



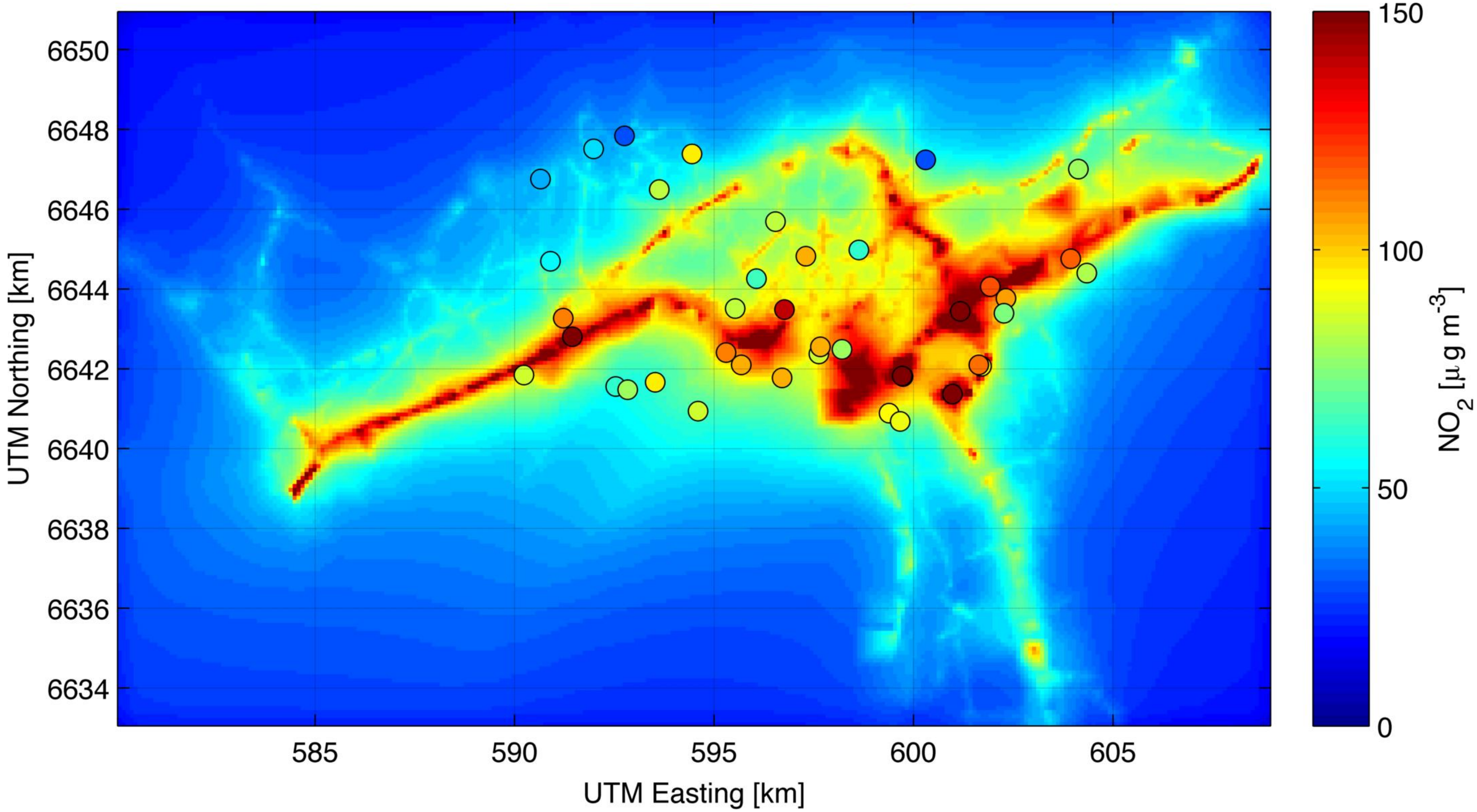
Synthetic observations of NO_2 concentrations generated over Oslo.

Oslo Data Fusion Example: Model information (auxiliary data)



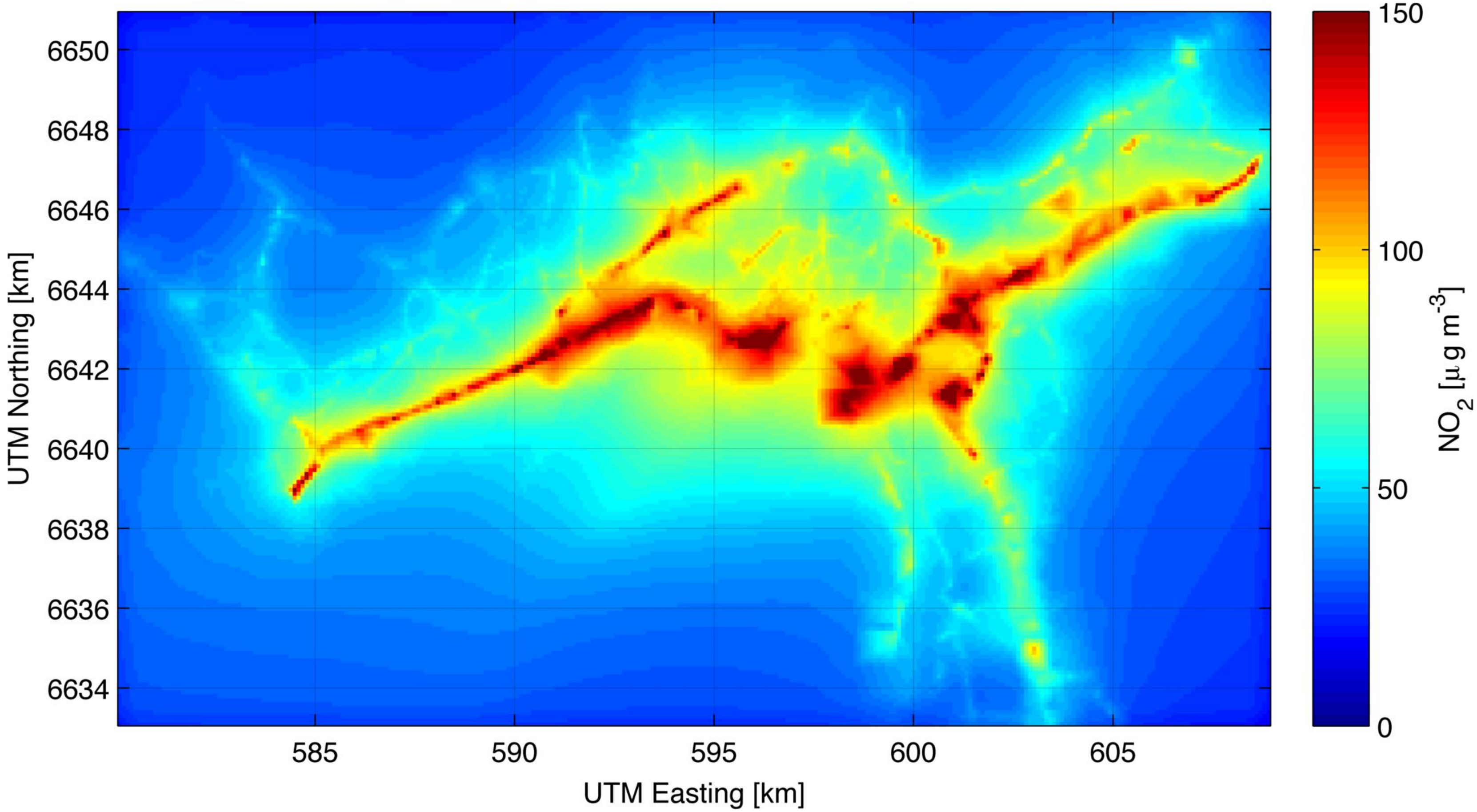
Average NO_x concentrations over the Oslo region for 2008 as provided by the EPISODE air pollution dispersion model (Slørdal et al 2008). Methodology for high-resolution model output developed by Bruce Denby at NILU.

Oslo Data Fusion Example: Model with observations



Model data (auxiliary information) and synthetic observations over Oslo. Note that the observations agree well with the model information in some areas but show significant discrepancies in other areas.

Oslo Data Fusion Example: Fused estimate



Fused product of NO2 concentrations over Oslo, combining both the information from the EPISODE dispersion model and the observations.

Data fusion implemented in R

- R provides a wealth of statistical libraries including Geostatistics
- Code for CITI-Sense implements the various steps in the data fusion methodology including log-transformation, regression, geostatistical interpolation, combination of maps, etc.
- Automated fitting of semivariogram model (possibility to fix some semivariogram parameters)
- Code is generic and not dependent on location, species etc.

Data fusion: Handling data gaps

- In addition to thorough quality control, bias correction etc., observations at CITI-SENSE nodes will be prone to frequent temporal data gaps due to sensor malfunction, outliers, etc.
- These gaps need to be filled in an objective and automated/operational fashion
- One solution: Predict missing values using Autoregressive Integrated Moving Average (ARIMA)
- Fully automated code has been developed in R to implement ARIMA-based gap-filling for CITI-SENSE nodes

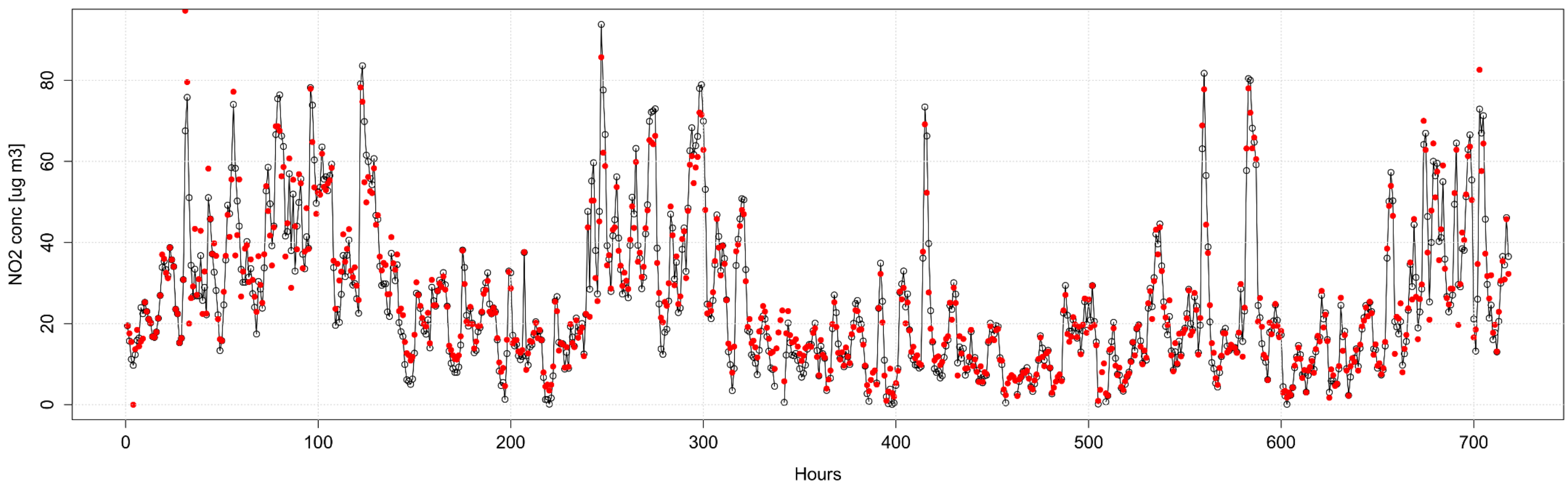
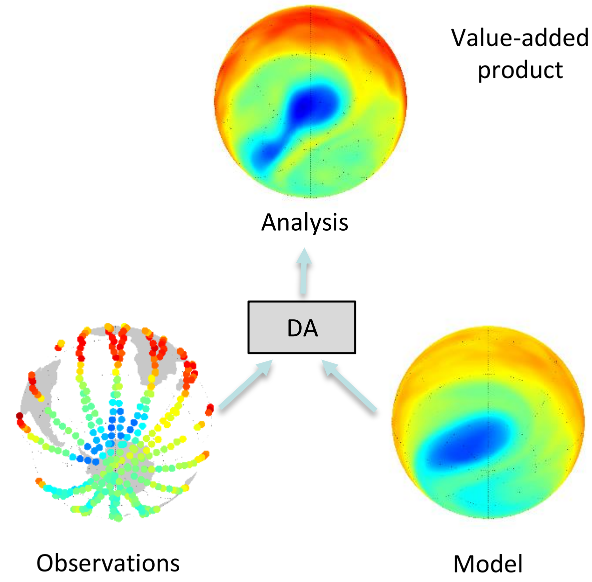


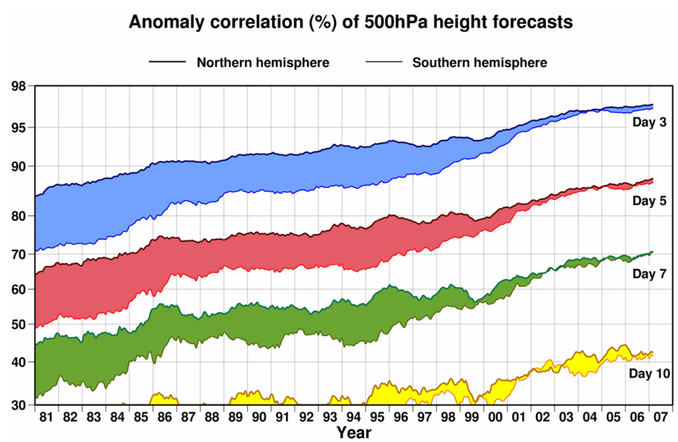
Figure illustrating the performance of ARIMA in predicting a missing value solely based on previously available information. The black markers show the true observations from an AQ station, the red points indicate the ARIMA-predicted value computed using only observations before that point in time.

Schneider et al. 2014: Spatial Data Fusion of Observations from Low-Cost AQ Sensors and Models for Urban-Scale Air Quality Mapping. COST TD1105 WG Meeting, 3-5 Dec 2014, Istanbul, Turkey

Next step: Data assimilation

- Combining observational and model information
- Based on mathematical principles (optimality; Bayes's rule)
- Added value:
  - Analysis better than either observations or model alone
  - Observations: filling in gaps
  - Model: constrain using observations (and learn about model deficiencies)
- Enormous success in NWP (ECMWF) and used for various elements of the Earth System
- Provides a way to account for the uncertainty of the observations (critical for "crowdsourced" microsensor data of AQ!)
- Data assimilation at the urban scale will be used experimentally within CITI-SENSE (as a demonstration for a relatively short period using past data)





Forecast improvement at EMCWF, based to a large extent on DA of satellite data

Next step: Data assimilation

- Assimilation of CITI-SENSE observation into a high-resolution air quality model, for example using the Ensemble Kalman Filter (EnKF)
- Cutting edge of research in DA: Has never been attempted at this spatial scale → only used in CITI-Sense for research, not operationally
- Advantages
  - Makes best use of information from both model and observations
  - Once set up can be run fully operationally
  - Based on vast amount of experience in operational NWP
- Disadvantages
  - Very complex to set up
  - Requires the data to be completely unbiased (problem with calibration drift of low-cost sensors)
  - Not used yet for urban scale AQ (uncharted terrain)

frontiers in ENVIRONMENTAL SCIENCE

REVIEW ARTICLE
published: 28 May 2014
doi: 10.3389/fenvs.2014.00016

Data assimilation: making sense of Earth Observation

William A. Lahoz* and Philipp Schneider

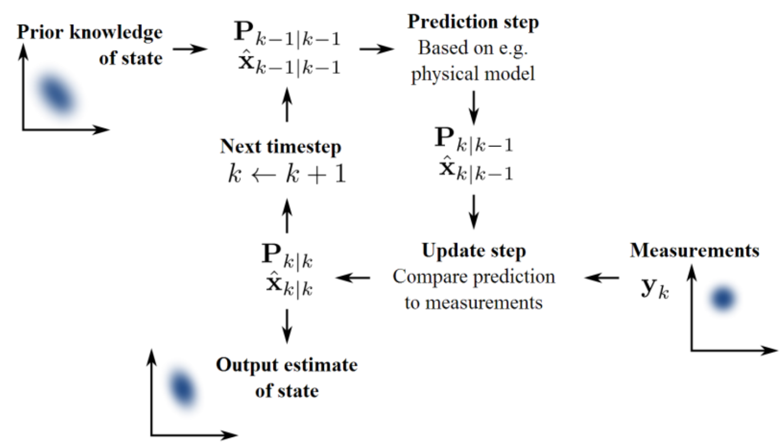
INBY, NILU – Norwegian Institute for Air Research, Kjeller, Norway

Edited by:
Annika Seppälä, Finnish Meteorological Institute, Finland

Reviewed by:
Avelino Florentino Arellano, University of Arizona, USA
Johanna Tamminen, Finnish Meteorological Institute, Finland

Climate change, air quality, and environmental degradation are important societal challenges for the Twenty-first Century. These challenges require an intelligent response from society, which in turn requires access to information about the Earth System. This information comes from observations and prior knowledge, the latter typically embodied in a model describing relationships between variables of the Earth System. Data assimilation provides an objective methodology to combine observational and model information to provide an estimate of the most likely state and its uncertainty for the whole Earth System.

Data assimilation review paper in *Frontiers in Environmental Science* (Lahoz WA and Schneider P (2014) Data assimilation: making sense of Earth Observation. *Front. Environ. Sci*. **2**:16. doi: 10.3389/fenvs.2014.00016)



Schematic showing the update loop of the Kalman Filter

Summary

- CITI-Sense requires an automated, operational, near-real-time system for mapping the observations onto a spatial grid
- Simple spatial interpolation is not able to provide realistic mapping results
- Data fusion of a spatial proxy (model) with the observations provides superior results
- Spatial proxy can be the output from land-use regression or a dispersion model
- In the end, this methodology provides a simple yet powerful automated technique for combining the information from both models and CITI-Sense observations in real time

Thank you for your attention!

Philipp Schneider

Email: ps@nilu.no

http://www.nilu.no