Measuring vs. Calculating: Why both models of determining fine-granular air quality might not even be wrong

Till Riedel, till.riedel@kit.edu
Matthias Budde, matthias.budde@kit.edu

KIT Department of Informatics

TECO

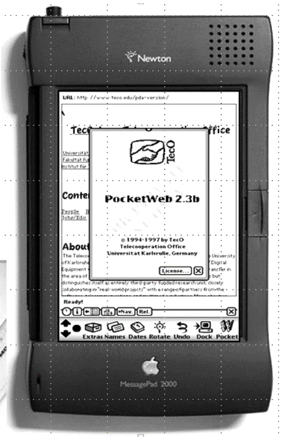
- Head: Prof. Michael Beigl
  Lab Lead: Dr. Till Riedel

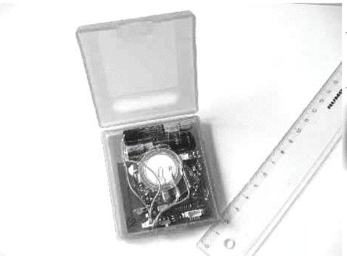
- Founded >20 years ago as cooperation between Digital Campus Engineering Center (then SAP CEC) and Uni Karlsruhe (now KIT)

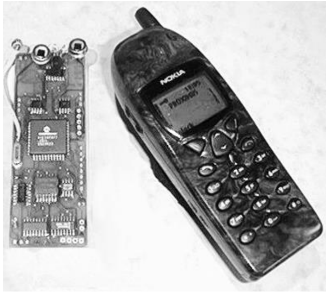
- Application oriented research in telematics

- 15 RAs, 20-30 Students
  - Started 100% 3rd party funded now part of Pervasive Computing Chair
  - EU-Projects, national funding
  - Industry, e.g. SAP, IBM, Huawei, Daimler, Bosch, Siemens, TRUMPF, Phillips, Infineon, KDDI

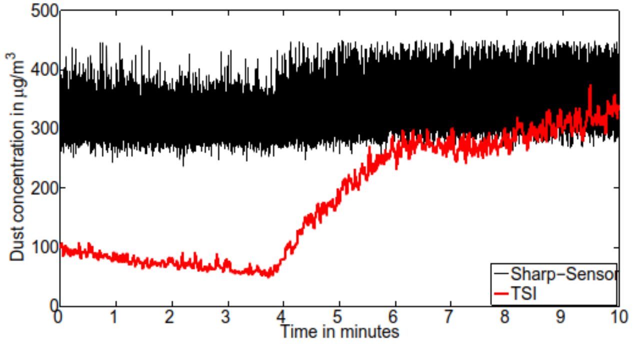
- Early Focus on Ubiquitous Computing (now IoT)
  - First mobile web browser: PocketWeb (1996)
  - First context aware mobile phone: TEA (1998)
  - First smart everyday artefact: MediaCup (1999)
  - Fist European Wireless Sensor Network Platform: SmartITs (2002)





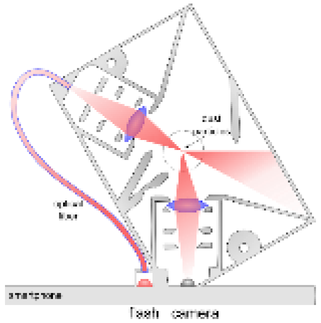


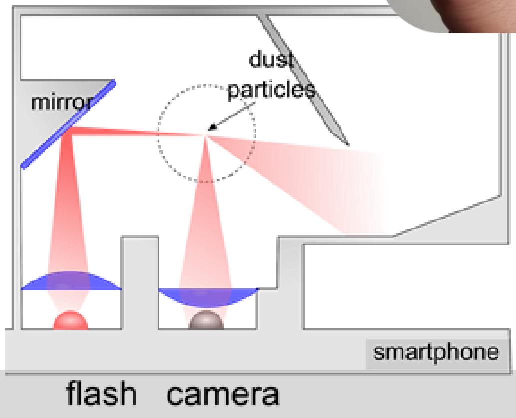
How it began 1/2: Low Cost Sensor (2011)



How it began 2/2: Smartphone Retrofit Sensor (2014)

- Idea: Clip-on PM sensor module for smartphones
- 4 generations of prototypes:
  - 3D printed for rapid prototyping
  - Light from flash is rerouted using an optical fiber respectively a mirror
  - Active versions with externally powered LEDs





Air quality networks are perfect basis for research in ubiquitous systems

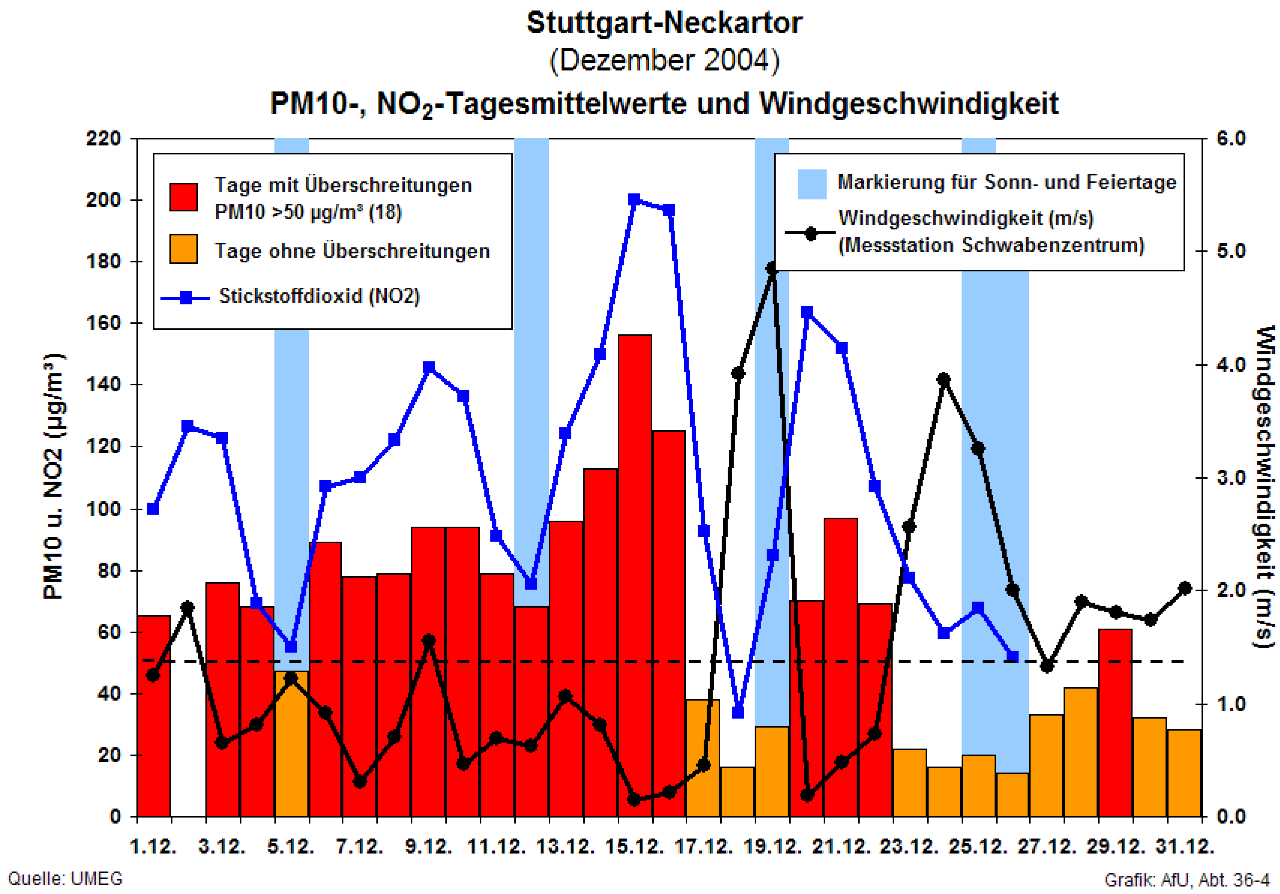
- Novel mobile and personal sensing methods…
- Usability and UX for cititizen scientists…
- Mobile calibration…
- Privacy issues when sharing data…
- New IoT networks…
- Interconnection with smart city control…
- Real world data sources needed for modeling…
- Data mining of real world data sources…

Air quality networks are perfect basis for research in ubiquitous systems

- Matthias Budde, Matthias Berning, Mathias Busse, Takashi Miyaki, Michael Beigl (2012) **Handheld Particulate Matter Measurements with COTS Sensors**, The 10th International Conference on Pervasive Computing (Pervasive 2012), pdf
- Matthias Budde, Matthias Berning, Mathias Busse, Takashi Miyaki, Michael Beigl (2012) The TECO Envboard: a Mobile Sensor Platform for Accurate Urban Sensing - and More, The 9th International Conference on Networked Sensing Systems, p. 1-2, , Best Demo Nominee, pdf, doi:10.1109/INSS.2012.6240573
- Matthias Budde, Mathias Busse, Michael Beigl (2012) Investigating the Use of Commodity Dust Sensors for the Embedded Measurement of Particulate Matter, The 9th International Conference on Networked Sensing Systems (INSS 2012), p. 1-4, , pdf, doi:10.1109/INSS.2012.6240545
- Matthias Budde, Pierre Barbera, Rayan El Masri, Till Riedel, Michael Beigl (2013) **Retrofitting Smartphones to be Used as Particulate Matter Dosimeters,** 17th International Symposium on Wearable Computers (ISWC'13), p. 139-140, pdf, doi:10.1145/2493988.2494342
- Matthias Budde, Rayan El Masri, Till Riedel, Michael Beigl (2013) Enabling Low-Cost Particulate Matter Measurement for Participatory Sensing Scenarios, 12th International Conference on Mobile and Ubiquitous Multimedia (MUM 2013), pdf, doi:10.1145/2541831.2541859
- Matthias Budde, Julio De Melo Borges, Stefan Tomov, Till Riedel, Michael Beigl (2014) Improving Participatory Urban Infrastructure Monitoring through Spatio-Temporal Analytics, 3rd ACM SIGKDD International Workshop on Urban Computing (UrbComp'14)
- Matthias Budde, Julio De Melo Borges, Stefan Tomov, Till Riedel, Michael Beigl (2014) Leveraging Spatio-Temporal Clustering for Participatory Urban Infrastructure Monitoring, The First International Conference on IoT in Urban Space (UrbIoT'14),
- Matthias Budde, Lin Zhang, Michael Beigl (2014) Distributed, low-cost particulate matter sensing: scenarios, challenges, approaches, ProScience 1(1st International Conference on Atmospheric Dust (DUST 2014)), p. 230-236, , pdf, doi:10.14644/dust.2014.038
- Matthias Budde, Lin Zhang, Michael Beigl (2014) Challenges and Approaches for Low-Cost Particulate Matter Sensing in Smart Cities, I International Conference on Atmospheric Dust – DUST 2014 3, p. 55, pdf
- Matthias Budde, Marcel Köpke, Michael Beigl (2015) Robust In-situ Data Reconstruction from Poisson Noise for Low-cost, Mobile, Non-expert Environmental Sensing, 19th International Symposium on Wearable Computers (ISWC'15), pdf, doi:10.1145/2802083.2808406
- Julio De Melo Borges, Matthias Budde, Oleg Peters, Till Riedel, Andrea Schankin, Michael Beigl (2016) EstaVis: A Real-World Interactive Platform for Crowdsourced Visual Urban Analytics, Proceedings of the Second International Conference on IoT in Urban Space - Urb-IoT'16
- Jan-Frederic Markert, Matthias Budde, Gregor Schindler, Markus Klug, Michael Beigl (2016) **Private Rendezvous-based Calibration of Low-Cost Sensors for Participatory Environmental Sensing**, 2nd EAI International Conference on IoT in Urban Space (UrbIoT'16),

Air quality networks are perfect basis for research in ubiquitous systems

- Matthias Budde, Michael Beigl (2016) Advances in Smartphone-based Fine Dust Sensing, II International Conference on Atmospheric Dust – DUST 2016 5, p. 23, pdf
- Julio De Melo Borges, Matthias Budde, Oleg Peters, Till Riedel, Michael Beigl (2016) Towards Two-Tier Citizen Sensing, 2nd IEEE International Smart Cities Conference (ISC2-2016), doi:10.1109/ISC2.2016.7580771
- Matthias Budde, Marcel Köpke, Michael Beigl (2016) Design of a Light-scattering Particle Sensor for Citizen Science Air Quality Monitoring with Smartphones: Tradeoffs and Experiences, ProScience 3(2nd International Conference on Atmospheric Dust – DUST2016), p. 13-20, url, doi:10.14644/dust.2016.003
- Jan-Frederic Markert, Matthias Budde, Gregor Schindler, Markus Klug, Michael Beigl (2018) Privacy-Preserving Collaborative Blind Macro-Calibration of Environmental Sensors in Participatory Sensing, EAI Endorsed Transactions on Internet of Things 18(10), pdf, doi:10.4108/eai.15-1-2018.153564
- Matthias Budde, Andrea Schankin, Julien Hoffmann, Marcel Danz, Till Riedel, Michael Beigl (2017) **Participatory Sensing or Participatory Nonsense? – Mitigating the Effect of Human Error on Data Quality in Citizen Science**. ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT) 1:3, doi: 10.1145/3131900
- Budde, M., Riedel, T., Beigl, M., Schäfer, K., Emeis, S., Cyrys, J., Schnelle-Kreise, J., Philipp, A., Ziegler, V., Grimm, H., Gratza, T.: *SmartAQnet – Remote and In-Situ Sensing of Urban Air Quality*, In: Remote Sensing of Clouds and the Atmosphere XXII, edited by Adolfo Comerón, Evgueni I. Kassianov, Klaus Schäfer, Richard H. Picard, Konradin Weber, Proceedings of SPIE, SPIE, Bellingham, WA, USA, Vol. 10424, 104240C-1 – 104240C-8 (2017); doi: 10.1117/12.2282698
- Budde, M., Riedel, T., Beigl, M., Riesterer, J., Schäfer, K., Emeis, S., Young, D., Cyrys, J., Schnelle-Kreis, J., Philipp, A., Petersen, E., Redelstein, J., Ziegler, V., Hank, M., Grimm, H., Hinterreiter, S., Gratza, T.: **SmartAQnet – high-resolution monitoring of urban air quality.** In: Proceedings of Abstract 11th International Conference on Air Quality, 172, Stick Air Quality Conference; 11th International Conference on Air Quality - Science and Application, 12. - 16. März 2018, Barcelona, Spanien; oral presentation
- Budde, M., Schäfer, K., Cyrys, J., Emeis, S., Gratza, T., Grimm, H., Hank, M, Hinterreiter, S., Petersen, E., Philipp, A., Redelstein, J., Riedel, T., Riesterer, J., Schnelle-Kreis, J., Young, D., Ziegler, V., Beigl, M.: *SmartAQnet – raum/zeitlich hochaufgelöste Erfassung der Luftqualität mit neuen Datenprodukten*. In: Umwelteinflüsse erfassen, simulieren und bewerten, 47. Jahrestagung der GUS 2018, Herausgeber Karl-Friedrich Ziegahn, Gesellschaft für Umweltsimulationen e.V., Pfinztal, Germany, 241-256; ISBN 978-9818507-2-7; 47. Jahrestagung der GUS, 21. - 23. März 2018, Stutensee-Blankenloch; oral presentation.



Stuttgart-Neckartor
(Dezember 2004)
PM10-, NO_2-Tagesmittelwerte und Windgeschwindigkeit
Tage mit Überschreitungen PM10 >50 µg/m³ (18)
Tage ohne Überschreitungen
Stickstoffdioxid (NO2)
Markierung für Sonn- und Feiertage
Windgeschwindigkeit (m/s) (Messstation Schwabenzentrum)
PM10 u. NO2 (µg/m³)
Windgeschwindigkeit (m/s)
0 20 40 60 80 100 120 140 160 180 200 220
0.0 1.0 2.0 3.0 4.0 5.0 6.0
1.12. 3.12. 5.12. 7.12. 9.12. 11.12. 13.12. 15.12. 17.12. 19.12. 21.12. 23.12. 25.12. 27.12. 29.12. 31.12.
Quelle: UMEG
Grafik: AfU, Abt. 36-4

STUTTGART
ERSTICKT

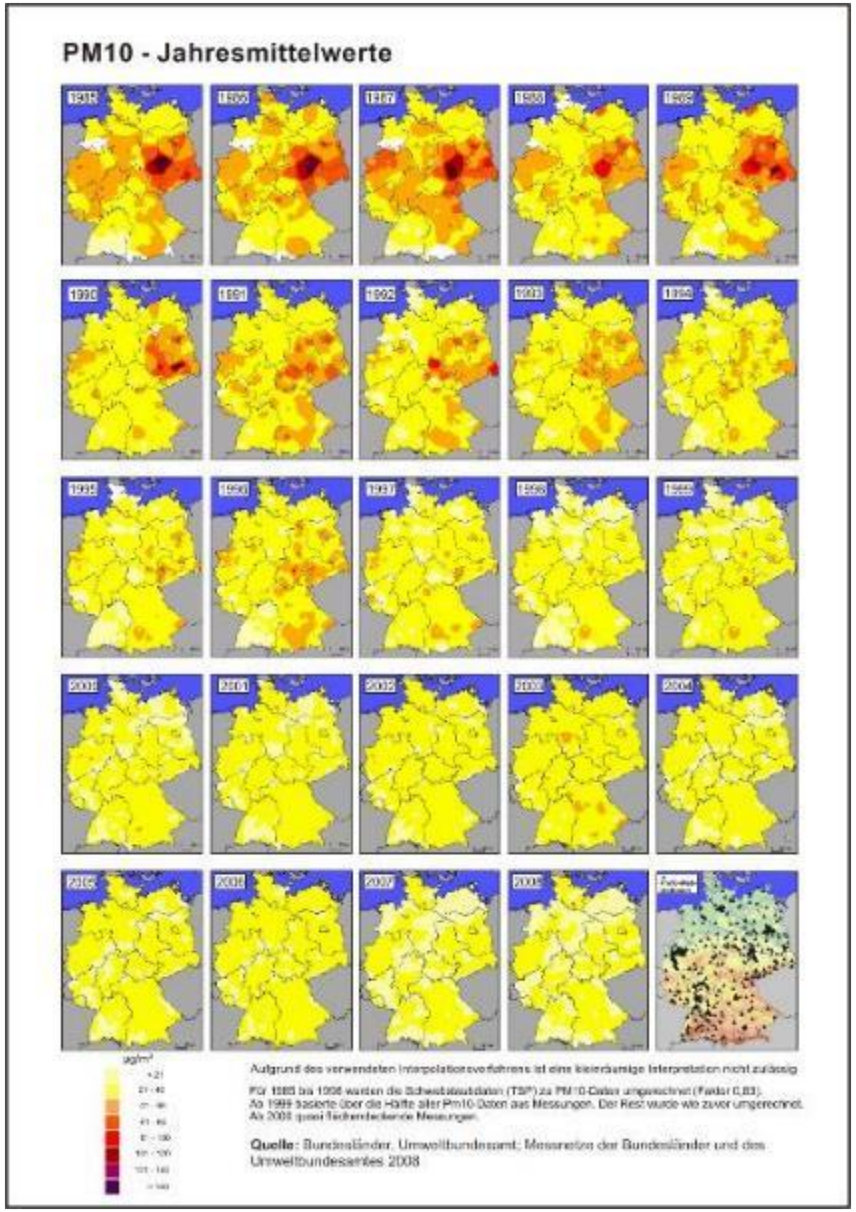
Is PM10 a problem after all?

- Research/regulation has triggered a positive development!

- Enforcement through measurement has lead to innovation and demand for new technologies!
  - Particle filters for non-diesel cars will come!
  - New problems: P25 → Ultrafine particles

- One problem:
  - local foreground concentrations can still be high even in Germany

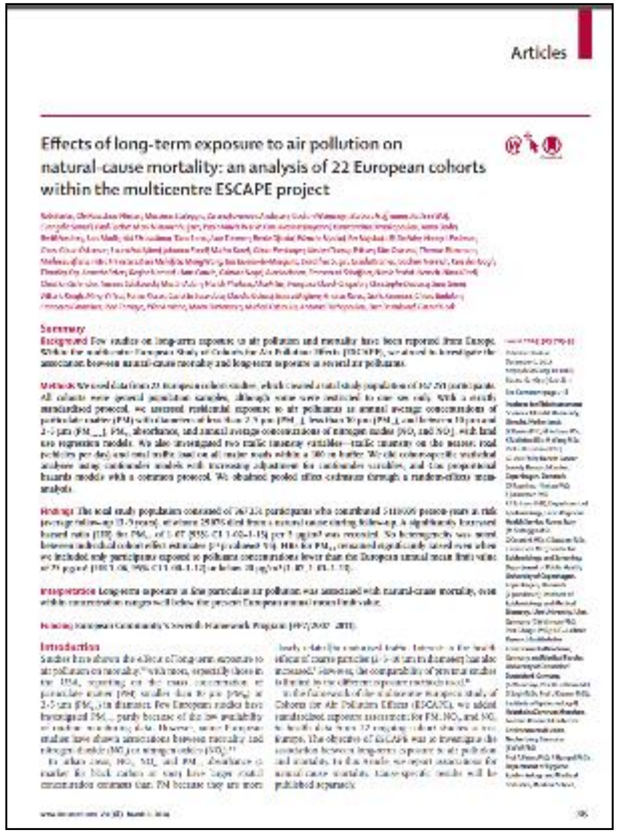
  - health effects of temporary personal exposure cannot be sufficiently quantified (→no regulation)



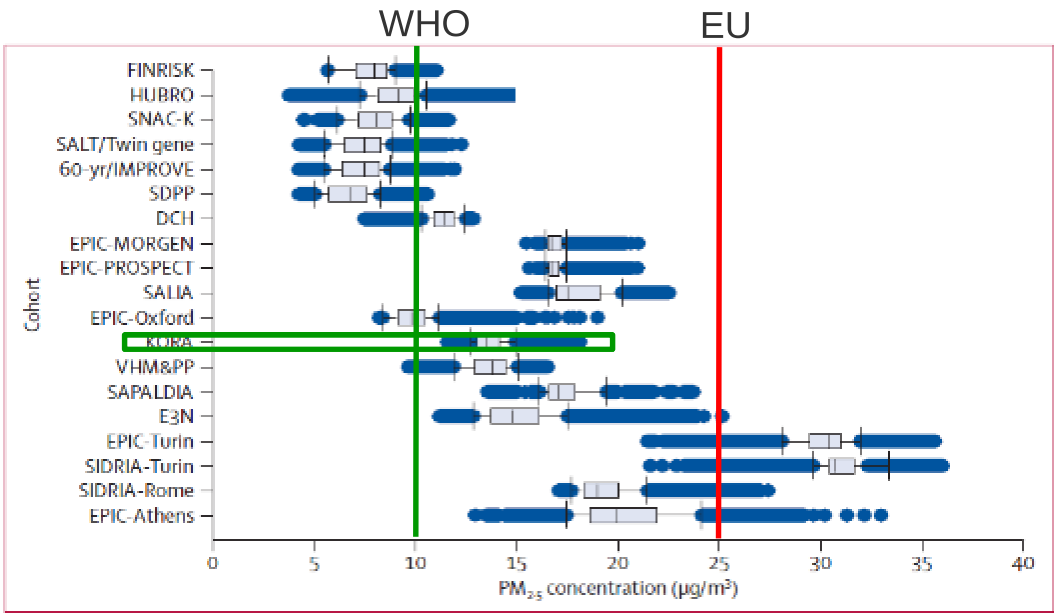
Danger: a matter of view point?

source:

Air Pollution and Mortality

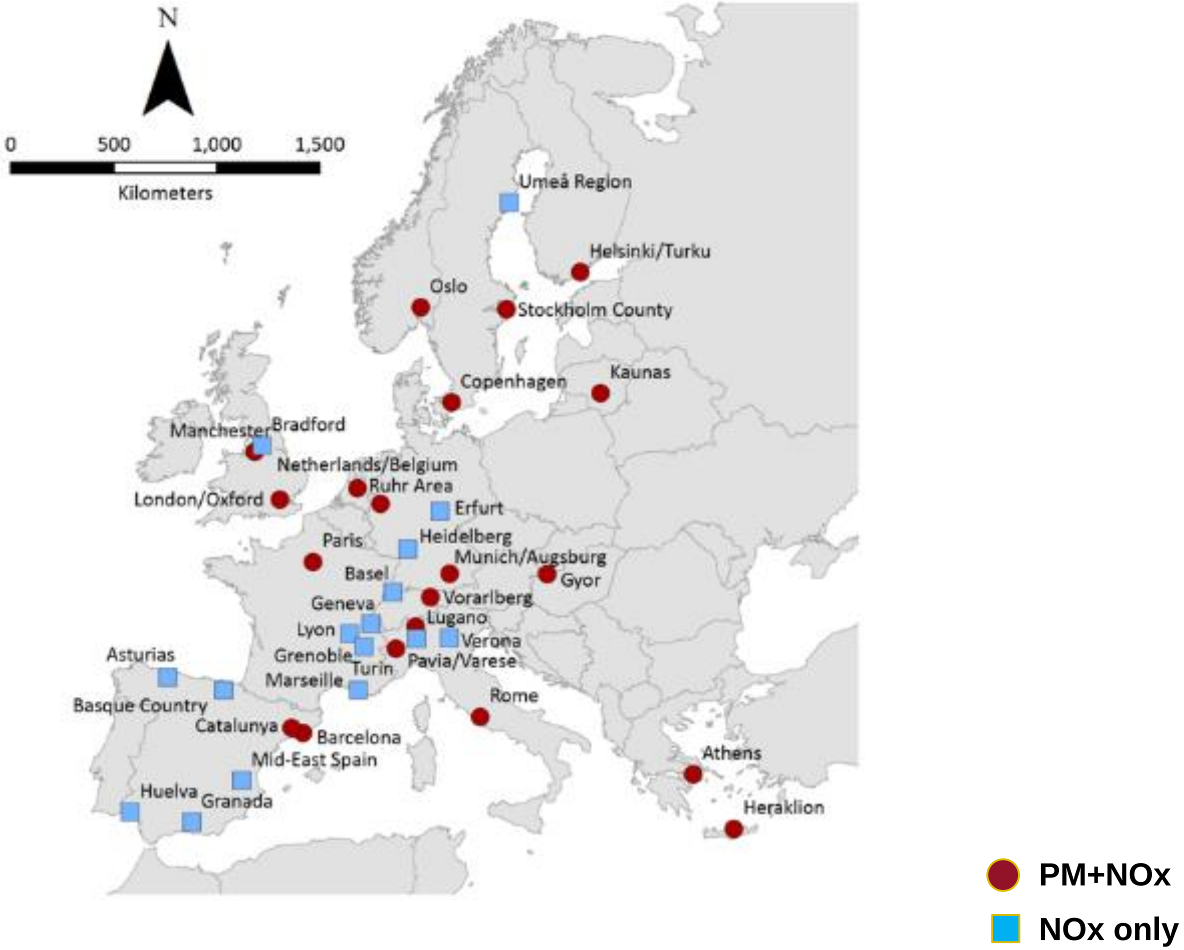

Articles

Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre ESCAPE project

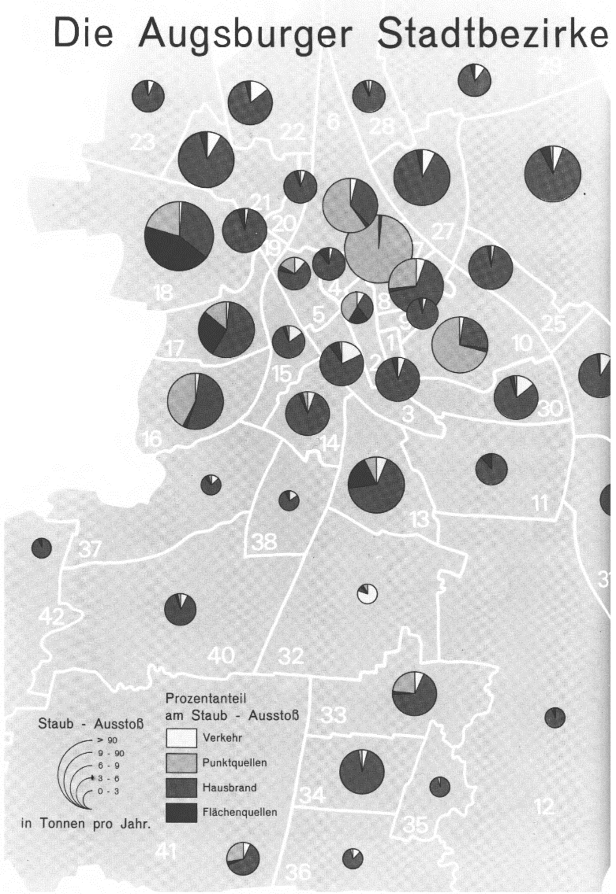


Beelen et al., Lancet 2014

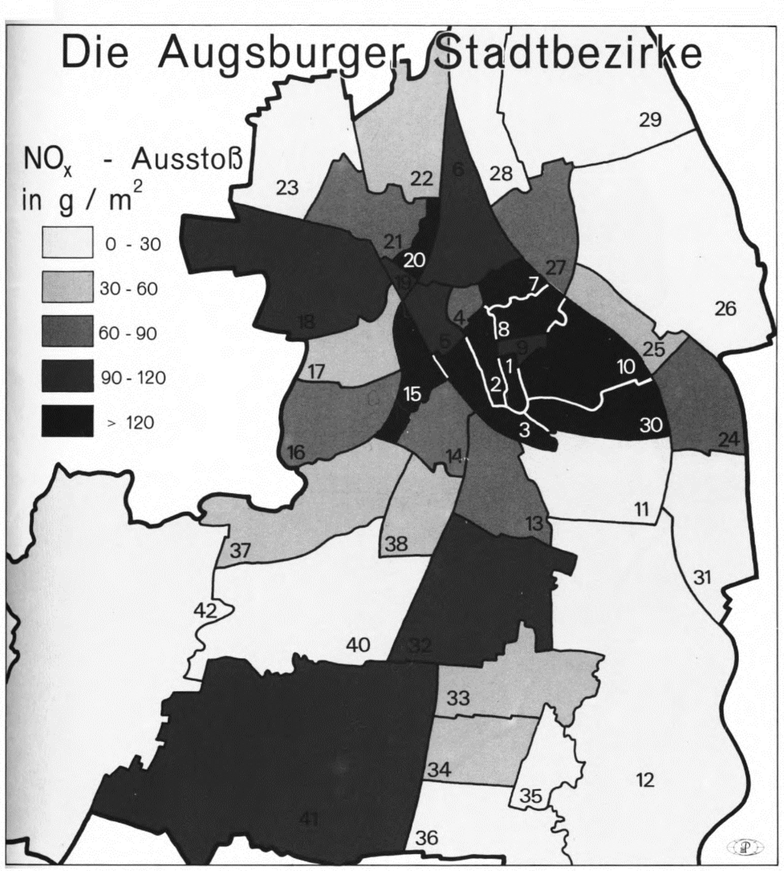
European Study of Cohorts for Air Pollution Effects : ESCAPE (2008 – 2013)



First epidmemiologic study in Augsburg: 1986



Staubemission



NOx - Ausstoß

1984 1995 1996 1999 2003 2008 2014 2015

Jacobeit J. (1986): Stadtklimatologie von Augsburg unter besonderer Berücksichtigung der lufthygienischen Situation sowie des Lärms. In Augsburger Geographische Hefte 6, 171p.

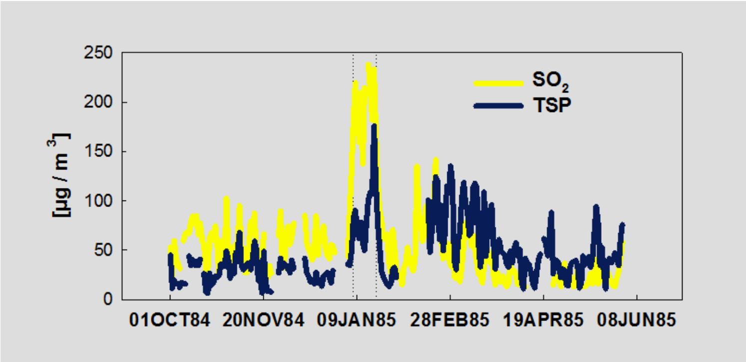
Milestone in Augsburg: Heart Attack Correlation to 1985 smog period

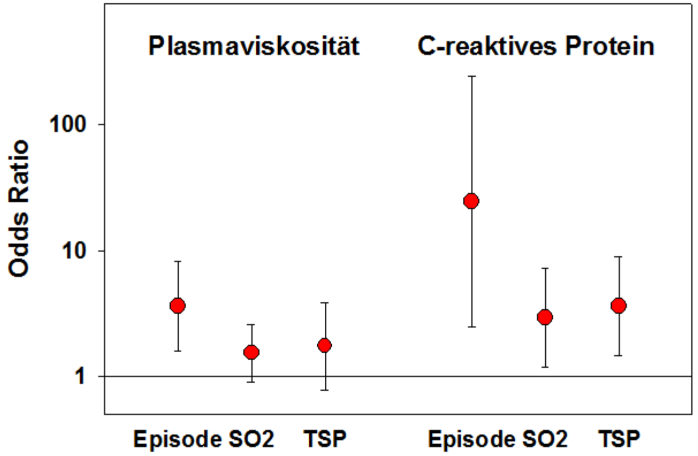
Karlsruher Institut für Technologie

THE LANCET

Increased plasma viscosity during an air pollution episode: a link to mortality?

Annette Peters, Angela Döring, H-Erich Wichmann, Wolfgang Koenig





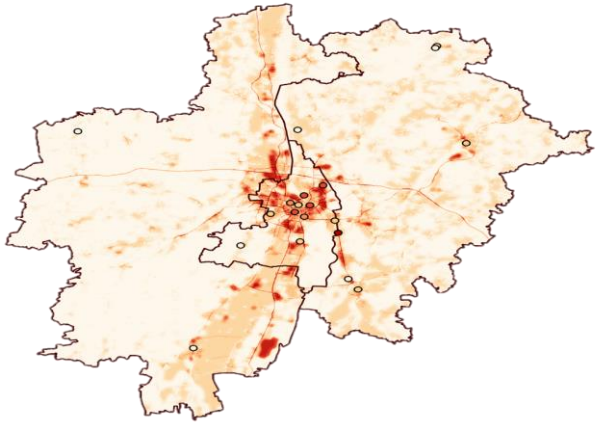
Während der Smogepisode 1985 konnte erstmalig eine systemische Reaktion nachgewiesen werden

Monica/Heart Attack Register

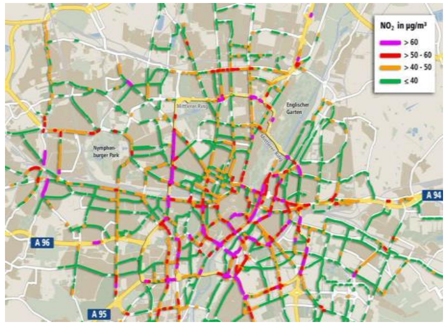
1984 1995 1996 1999 2003 2008 2014 2015

Peters et al., Lancet 1997

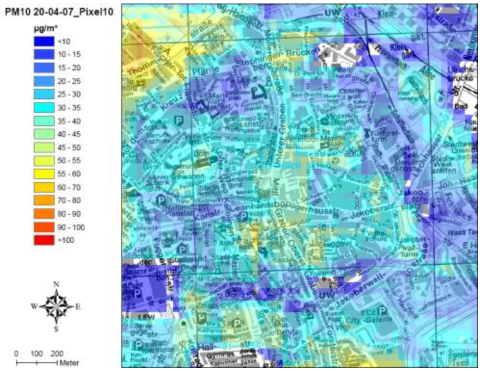
How to quantify exposure?



Source:helmholtz-muenchen.de



Source:merkur.de



Source:icaros net

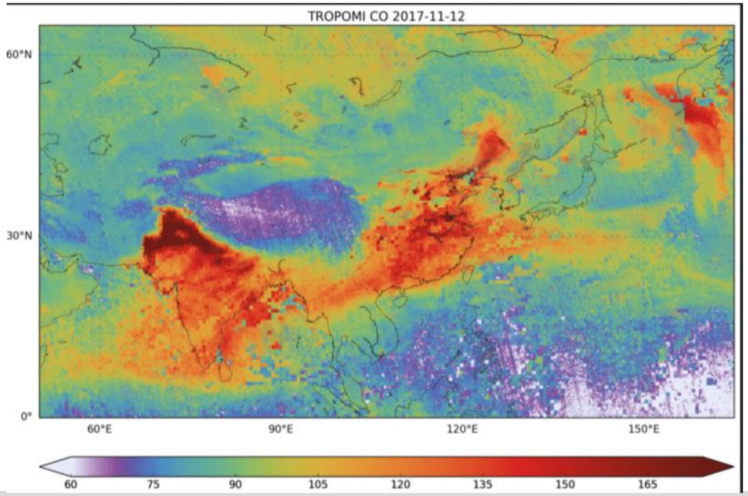
What is „correct“ local value?
People mostly not close to measurements.
People cannot carry equipment all the time…

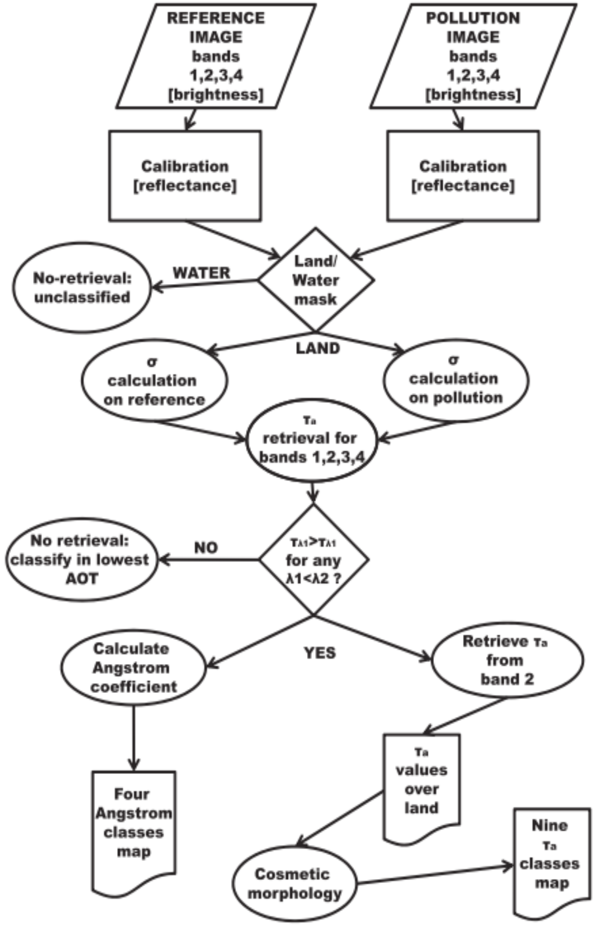
We need to interpolate/extrapolate →
Do we know what we are doing?

We can use satelite images to analyze the optical thickness!

Poluted image – Reference image → regression problem!

But how can we be sure we measure the concentration at 1.5m???





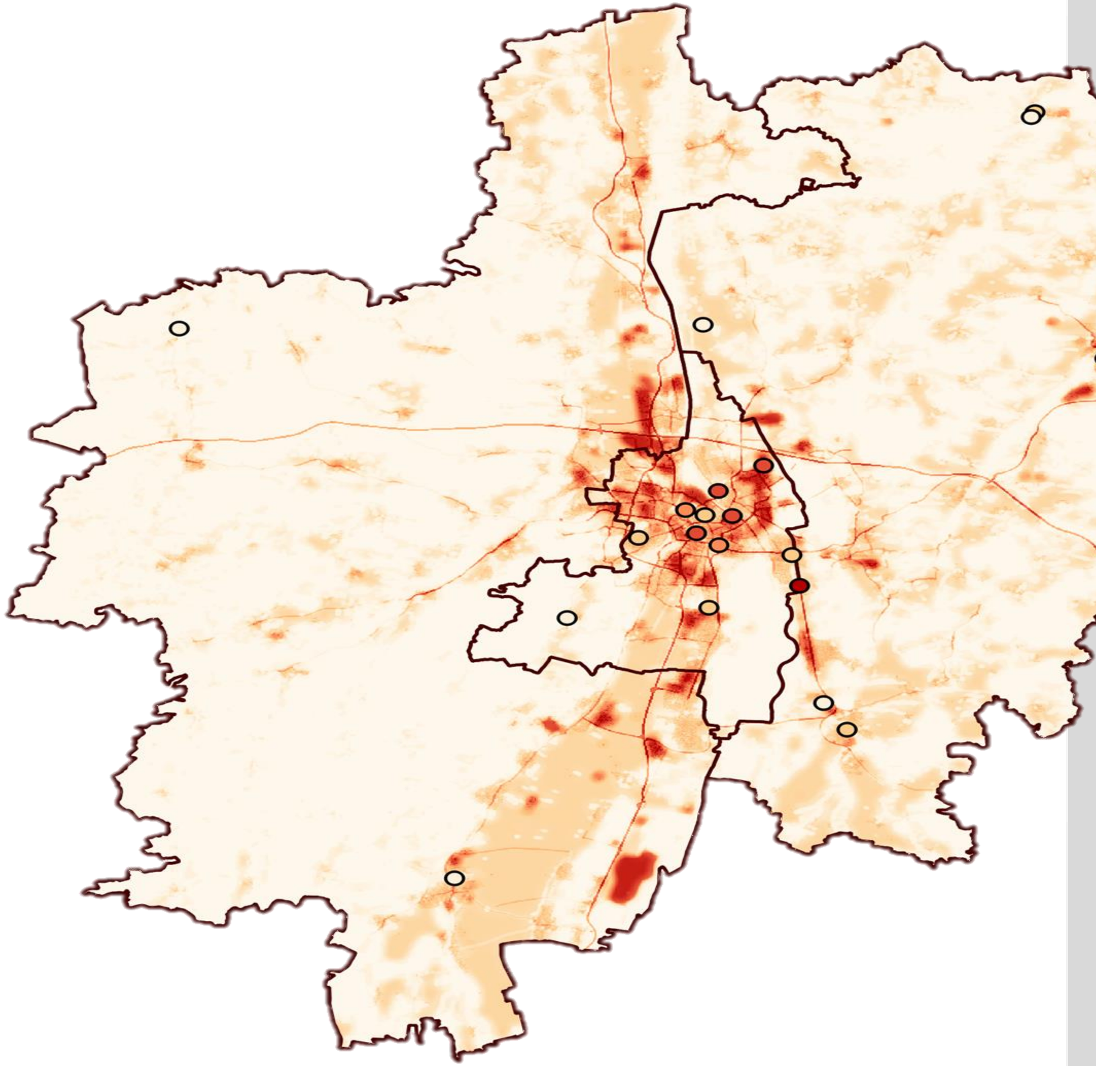
CHRISTINE Code for High ResolutIon Satellite mapping of optical ThIckness and ÅNgstrom Exponent. Part I: Algorithm and code

Nicolas I. Sifakis [a,*], Christos Iossifidis [b]

Beyond KORA ….

Continue Cooperative health research in Augsburg

- Continue important findings on health risks:
  - Eg. Correlation of diabetes with PM was recently proven
- In past using rather static models of PM based on land-use
  - 20 measurement points short-term
  - 5 long-term
- Can we better **estimate the personal exposure** to understand effects?
- Can we **create applications that proactively support health**?
- What is the role of measurements in **fine granular planning tools**?



Land Use Regression (LUR) Modeling

Land-use & land-cover variables:

1. Roadways:
   - Freeways(AADT)$_{1200}$
   - Major Arteries$_{500}$
   - Arteries$_{350}$
   - Streets(POP)$_{800}$
2. Railroads$_{250}$
3. Area under tree canopy$_{400}$
4. Elevation
5. X (Longitude)

$$
\begin{aligned}
\mathbf{NO_2} = {} & 7.7 + 1.1\times10^{-8} * FWY\_AADT_{1200} \\
& + 6.5\times10^{-4} * MAJ\_ART_{500} \\
& + 1.7\times10^{-3} * ARTERIES_{350} \\
& + 1.8\times10^{-8} (STREETS) * POP_{800} \\
& + 1.0\times10^{-3} * RAIL_{250} \\
& - 5.7\times10^{-6} * TREES_{400} \\
& - 1.0\times10^{-[illegible]} * ELEV + 1.4\times10^{-5} * (ELEV)^2 \\
& + 1.1\times10^{-4} * X\_DIST
\end{aligned}
$$

Source: Toward Breathable Cities For All: Linking Air Pollution, Vulnerable Populations and Human Health, Arbor Day Foundation

Does this mean if we plant enough trees we get clean air???

remote sensing

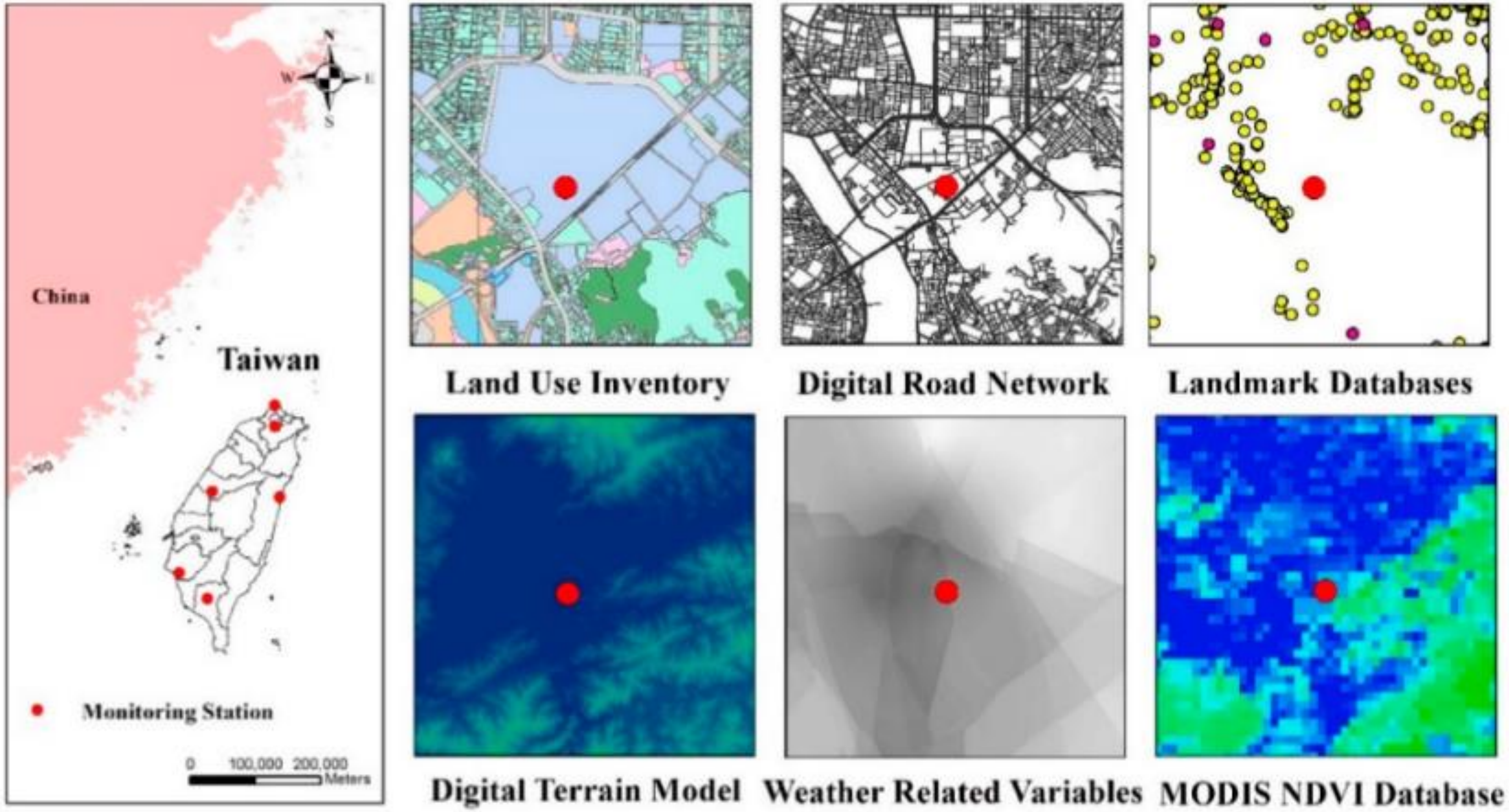
MDPI

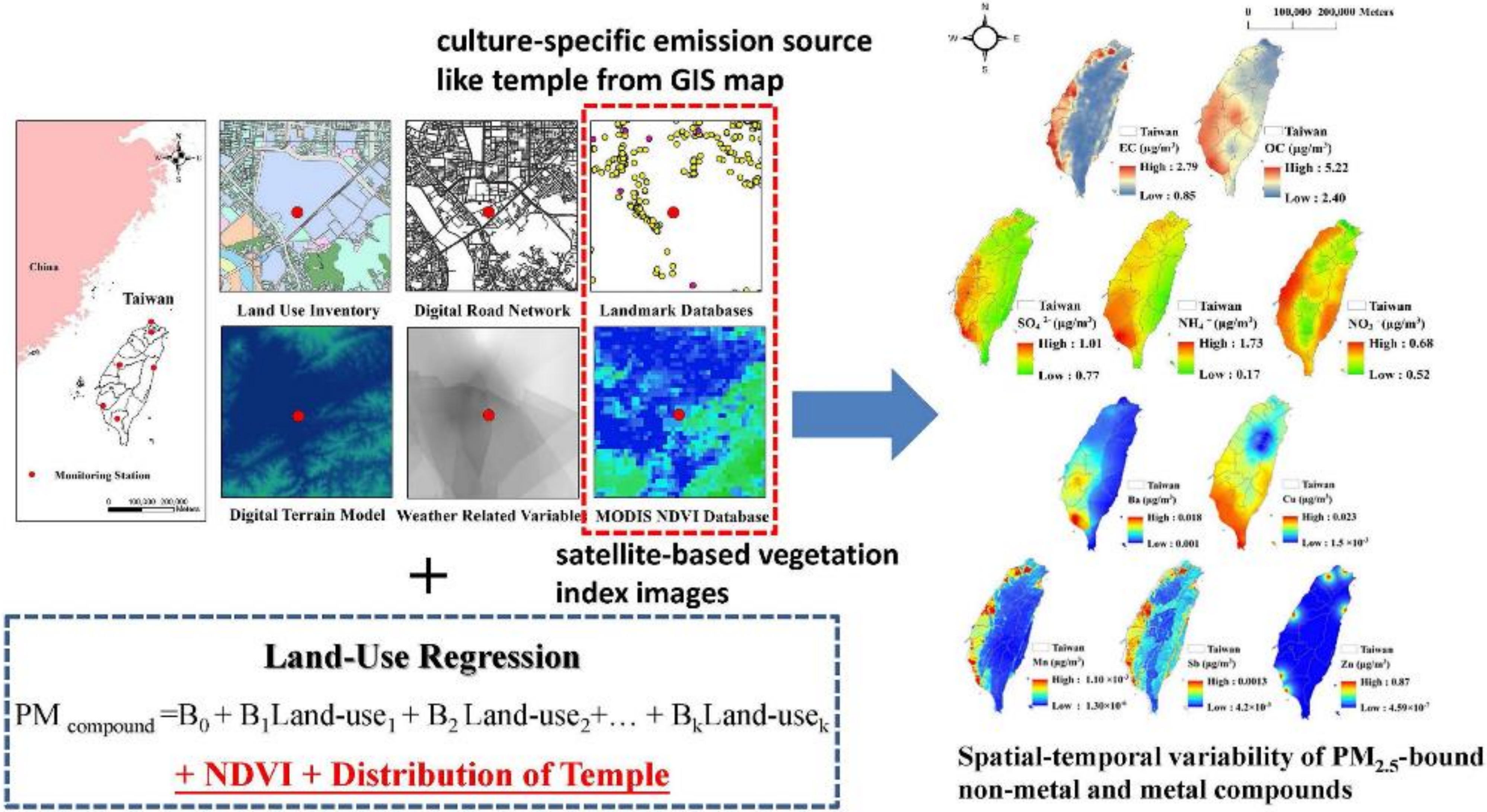
Article

Developing Land-Use Regression Models to Estimate $PM_{2.5}$-Bound Compound Concentrations

Chin-Yu Hsu [1], Chih-Da Wu [2,3,*], Ya-Ping Hsiao [1,2], Yu-Cheng Chen [1,4], Mu-Jean Chen [1] and Shih-Chun Candice Lung [5,6,7,*]

1 National Institute of Environmental Health Sciences National Health Research Institutes, Miaoli 350, Taiwan; gracecyhsu@nhri.org.tw (C.-Y.H.); hsiaoyapiau@gmail.com (Y.-P.H.); yucheng@nhri.org.tw (Y.-C.C.); zeromagi@nhri.org.tw (M.-J.C.)
2 Department of Forestry and Natural Resources, National Chiayi University, Chiayi 600, Taiwan
3 Department of Geomatics, National Cheng Kung University, Tainan 701, Taiwan
4 Department of Occupational Safety and Health, China Medical University, Taichung 404, Taiwan
5 Research Center for Environmental Changes, Academia Sinica, Taipei 115, Taiwan
6 Department of Atmospheric Sciences, National Taiwan University, Taipei 106, Taiwan
7 Institute of Environmental Health, School of Public Health, National Taiwan University, Taipei 100, Taiwan
* Correspondence: chidawu@mail.ncku.edu.tw (C.-D.W.); sclung@rcec.sinica.edu.tw (S.-C.C.L.);





Spatial-temporal variability of $PM_{2.5}$-bound non-metal and metal compounds

Machine Learning can use all correlations…

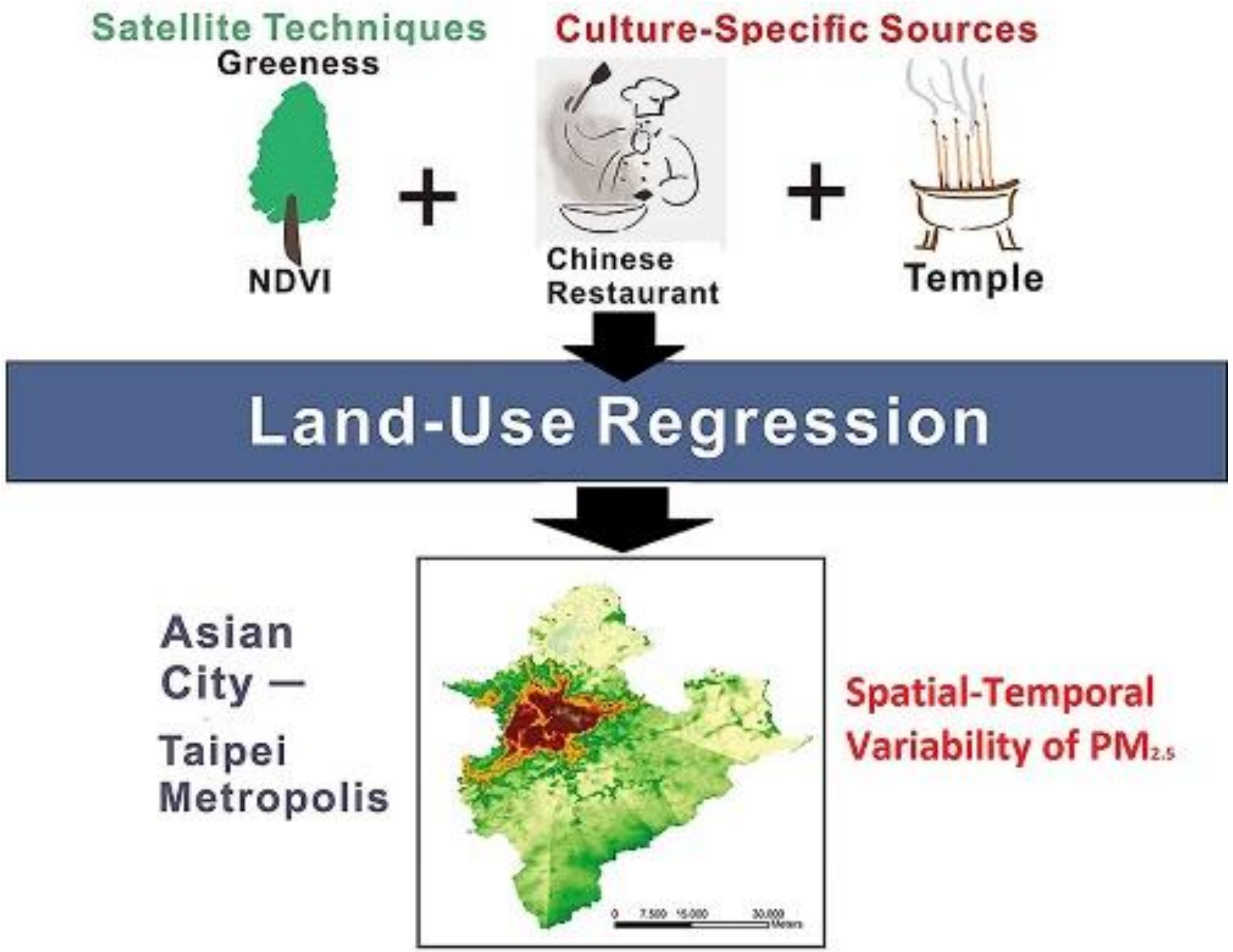
Greenness (r^2: −0.71 to −0.77)
Temples (r^2: 0.52 to 0.66)
Restaurants (r^2: 0.31 to 0.44)

More input variables?

More complicated models?

How locally overfitted can a model be?

→ More Measurements needed!!!



Wu, et. Al Land-use regression with long-term satellite-based greenness index and culture-specific sources to model $PM_{2.5}$ spatial-temporal variability, Environmental Pollution, Volume 224

Karlsruher Institut für Technologie

Low Cost Participatory Sensing of PM

Matthias Budde, matthias.budde@kit.edu

KIT Department of Informatics


UNA


Thermo
SCIENTIFIC

swerte
erschreitungen
ssungen

rtungen

ur Luftqualität

en Immissionsschutz

Immissionsdaten aus Baden-Württemberg

30.09.2018 08:00, vorläufige Werte

Luftschadstoffe

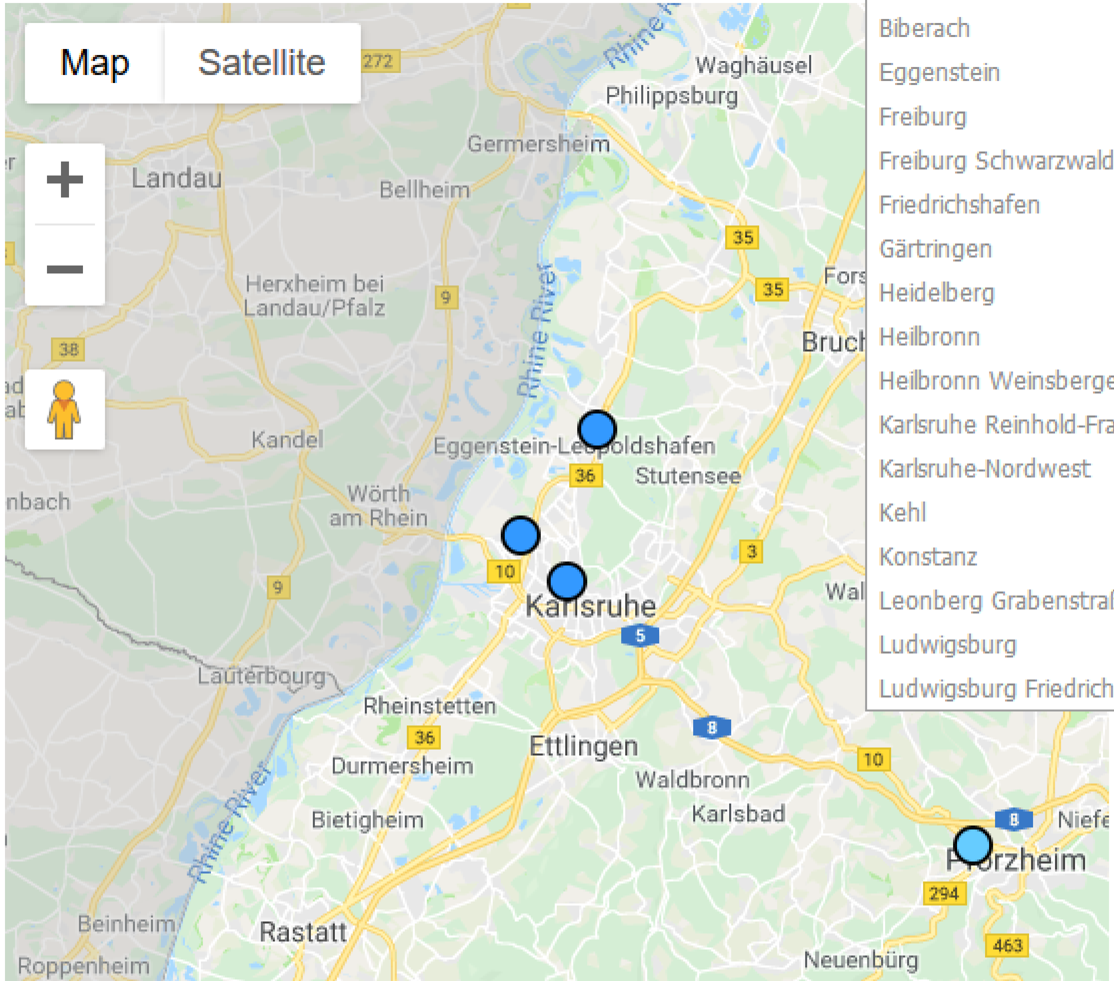
Feinstaub PM2,5 (kontinuierlich)

Messstellenauswahl

Bitte wählen

- Bitte wählen
- Aalen
- Baden-Baden
- Bernhausen
- Biberach
- Eggenstein
- Freiburg
- Freiburg Schwarzwaldstraße
- Friedrichshafen
- Gärtringen
- Heidelberg
- Heilbronn
- Heilbronn Weinsberger Straße-Ost
- Karlsruhe Reinhold-Frank-Straße
- Karlsruhe-Nordwest
- Kehl
- Konstanz
- Leonberg Grabenstraße
- Ludwigsburg
- Ludwigsburg Friedrichstraße

Karte Diagramm Tabelle



Range of measurement technology

Standard (gravimetric)	Professional (mostly optical)	Consumer (mostly optical)	Low-Cost (optical)

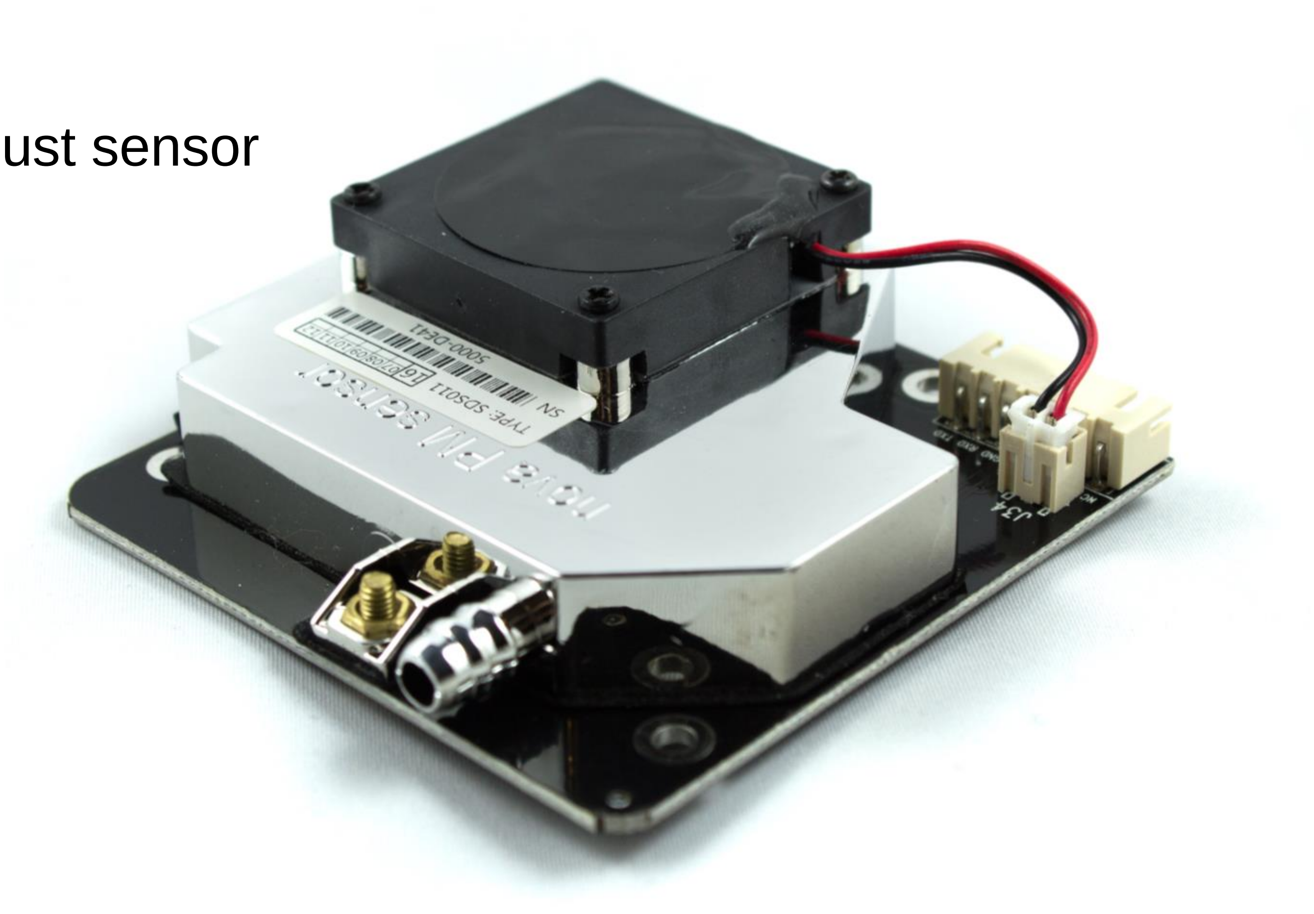
accurate	↔	noisy
stable	↔	unstable
expensive	↔	cheap
low temporal resolution	↔	high temporal resolution
delayed	↔	low latency
large	↔	small
static	↔	mobile

Tradeoff

Budde, M., Zhang, L., & Beigl, M. (2014). Distributed, low-cost particulate matter sensing: scenarios, challenges, approaches. In First International Conference on Atmospheric Dust (DUST 2014). Digilabs. https://doi.org/10.14644/dust.2014.038

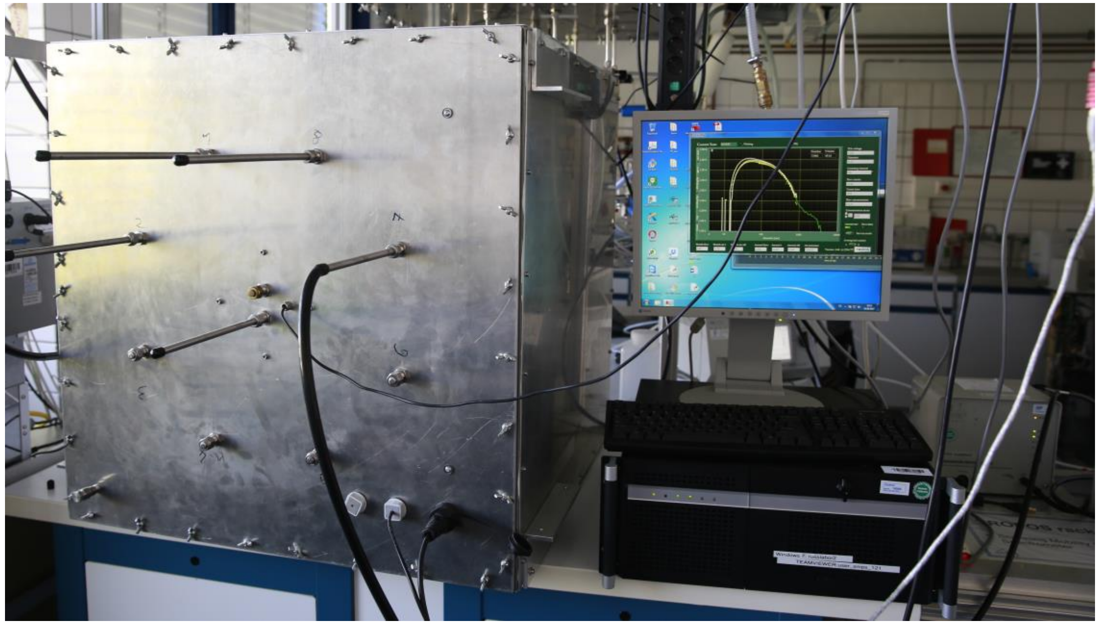
Alternative?: Low-cost Sensing?

- Light- or Laser-scattering
- Sensor cost 10-20 €
- Most common today for PM: SDS011 dust sensor

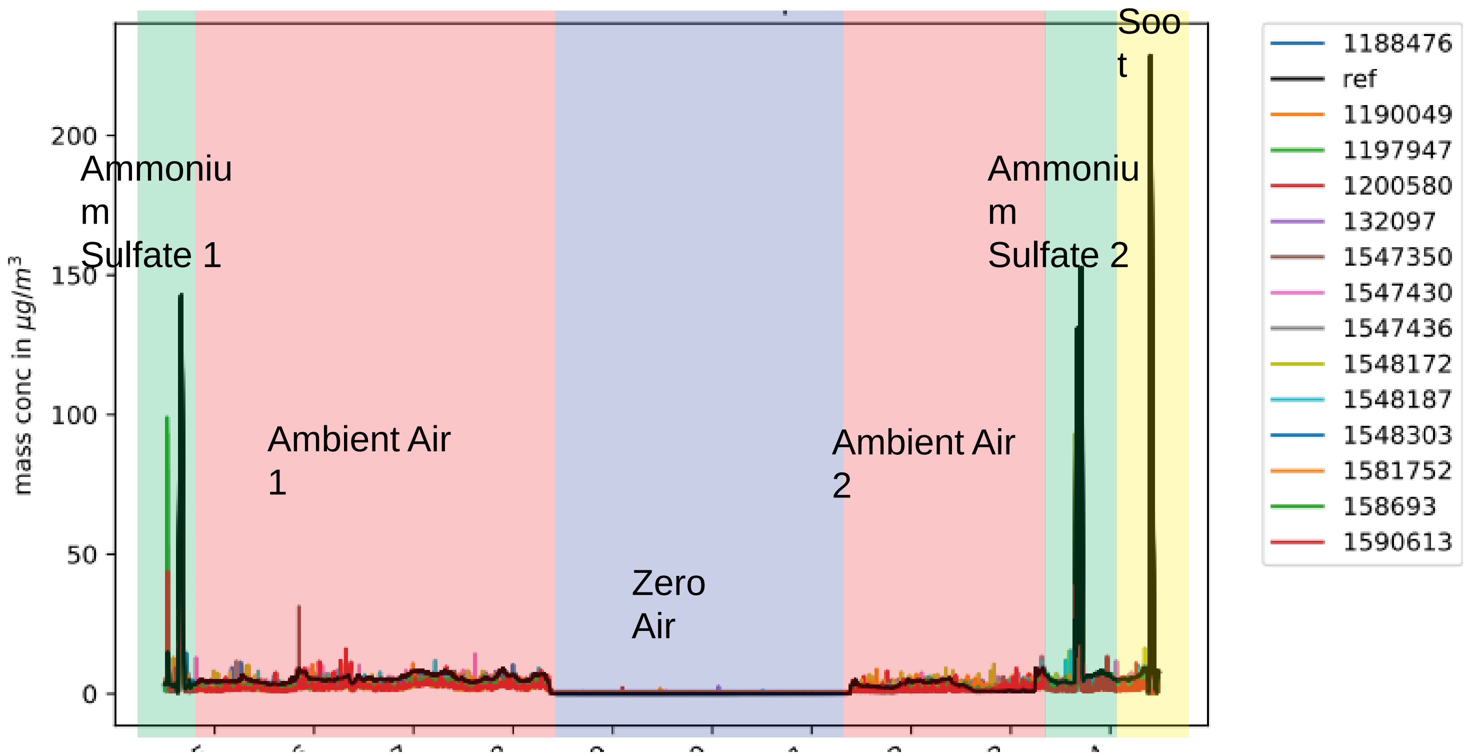


Polydisperse Particles: Setup

- Measurement at TROPOS, Leipzig, Germany
- 17 SDS011 sensors (14 with valid data)
- SMPS/APS reference with 92 aerodynamic channels

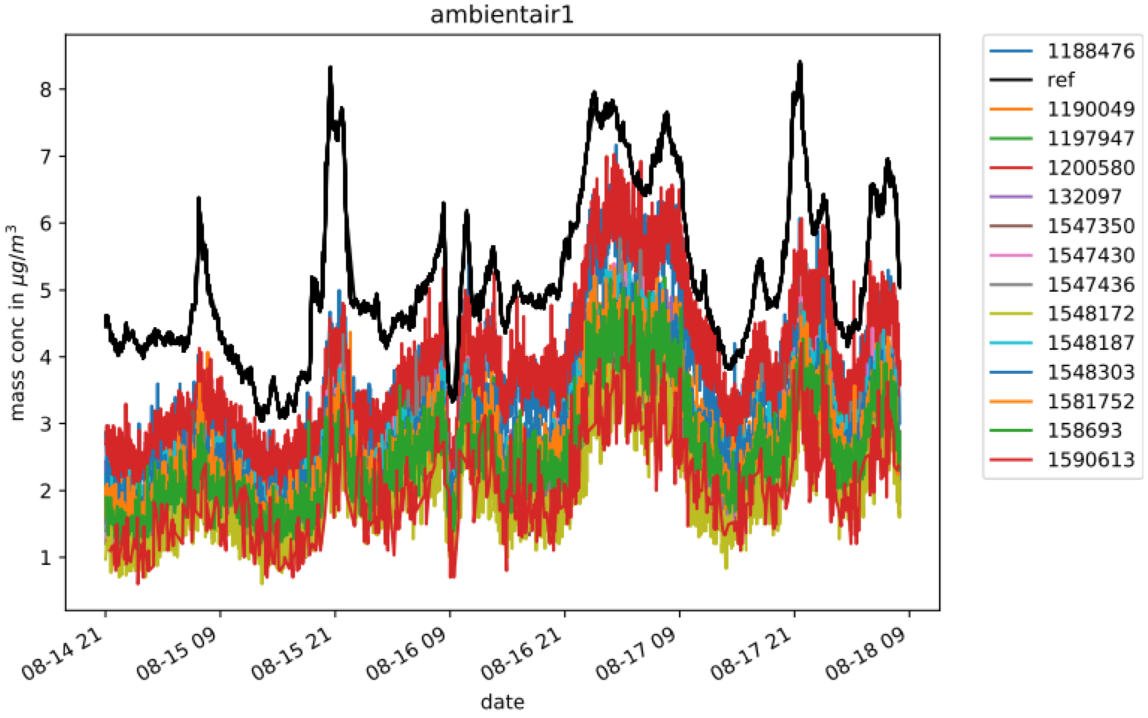


Polydisperse Particles: Experiments



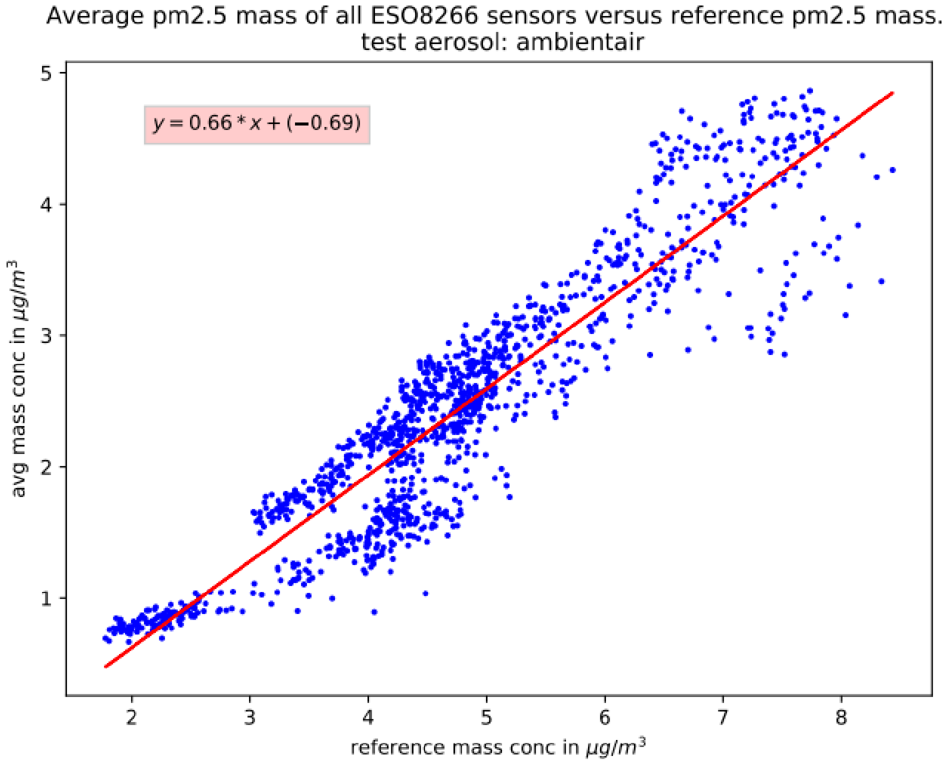
PM 2.5 Time Series

- Dynamics are generally captured well
- Readings have an offset
- Offset differs for individual sensors



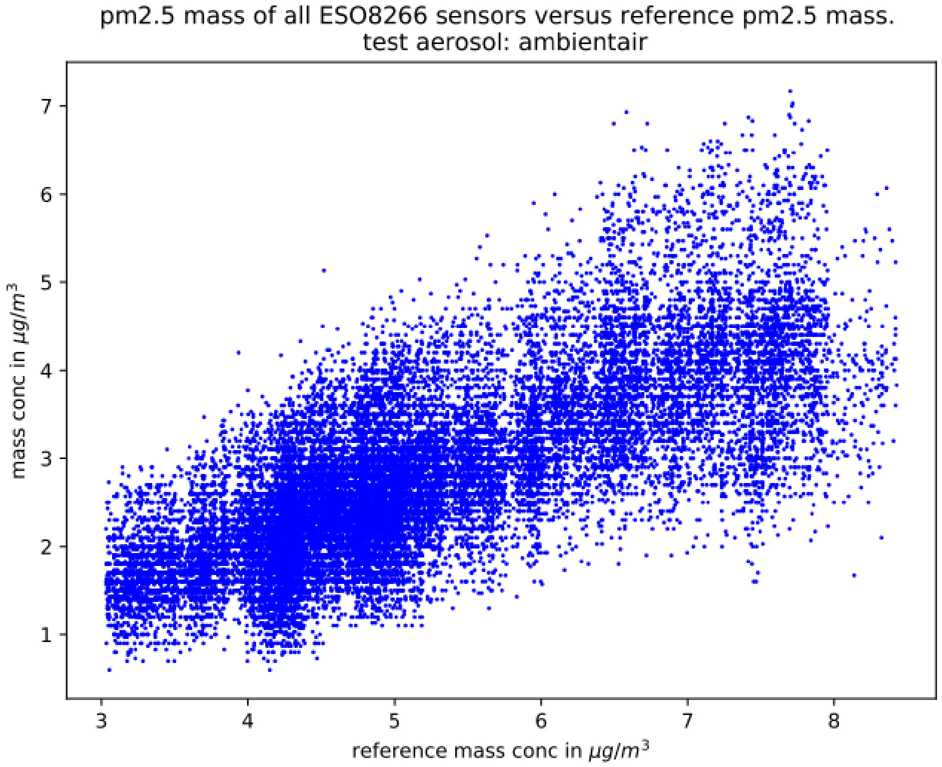
PM 2.5 Scatter Plots

- Good linearity
- Systematic misestimation of concentration
- Ambient air: on average 66% of the PM2.5 reference



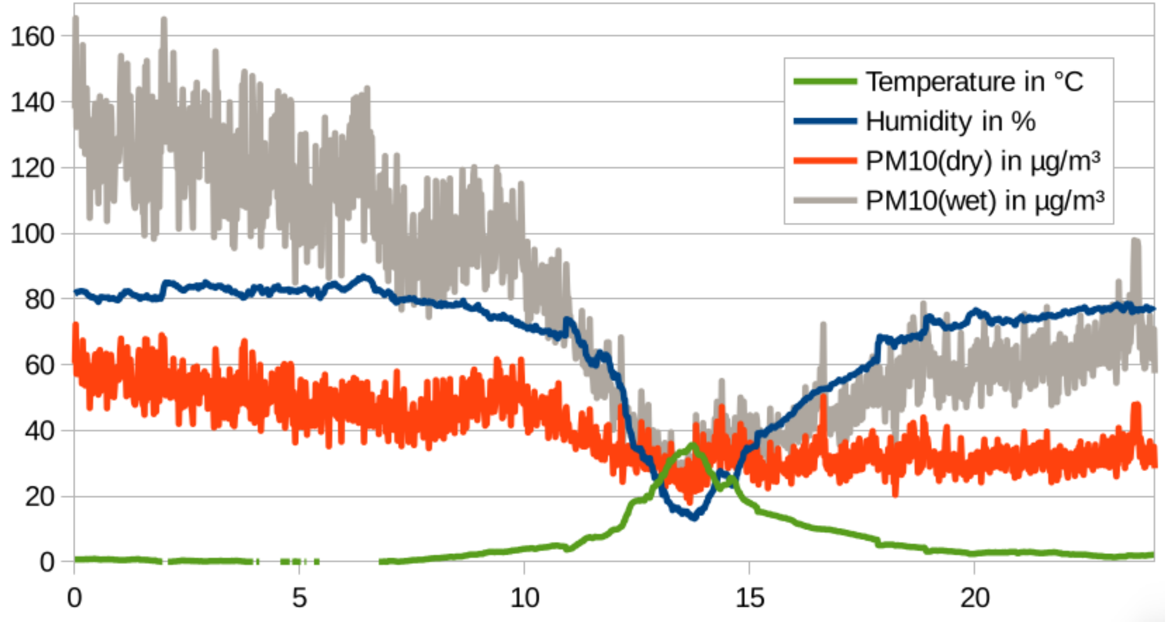
PM 2.5 Scatter Plots

- High variance between individual sensors
- Ranging from 45% to 85%



Real World Data: Humidity

- SDS011 only defined up to 70% RH
- Strong humidity dependence
- Fog is misread as fine dust



Conclusions?

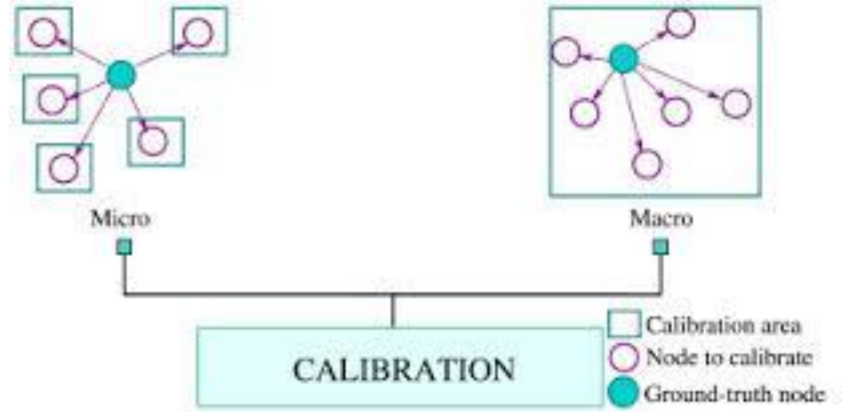
- SDS011 can capture dynamics with high temporal resolution
- Notable variance between sensors
- Humidity is a problem
- Suitability depends on application / further measures

Calibration

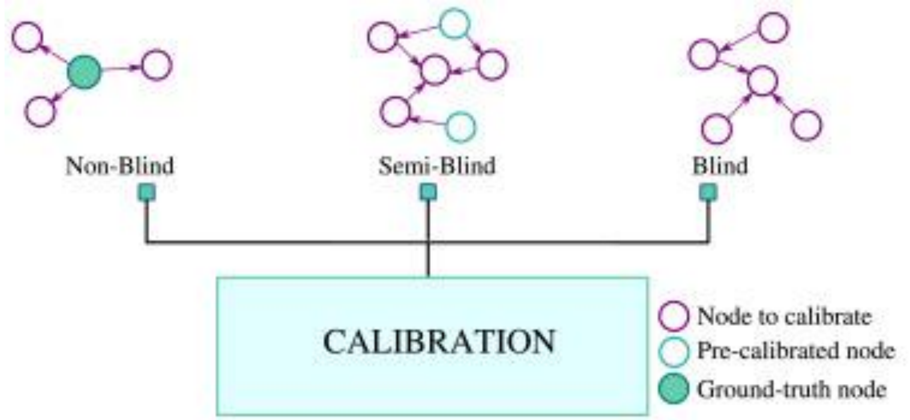
- Two basic approaches:
- Exposure to defined and well known concentrations and environmental conditions (mostly under lab conditions)
- Co-location with standard / high-precision reference (in field)
-

Some calibration „flavors“

- Micro vs macro

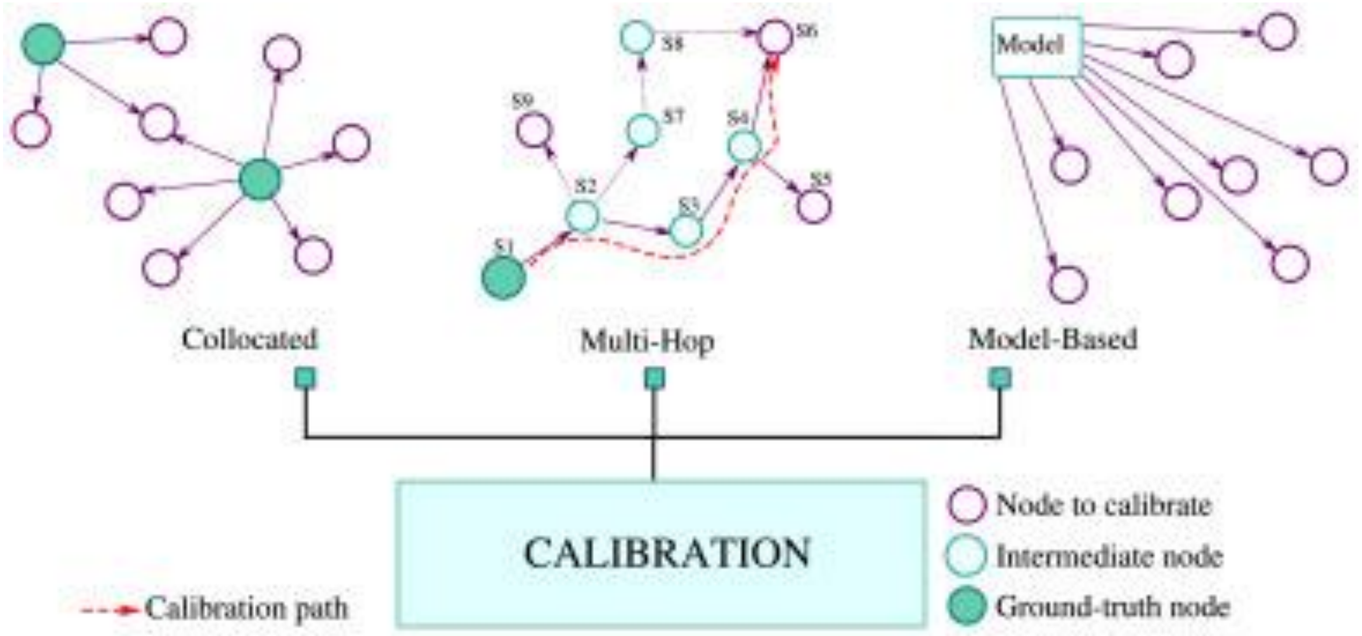


- Non-blind vs semi-blind vs blind

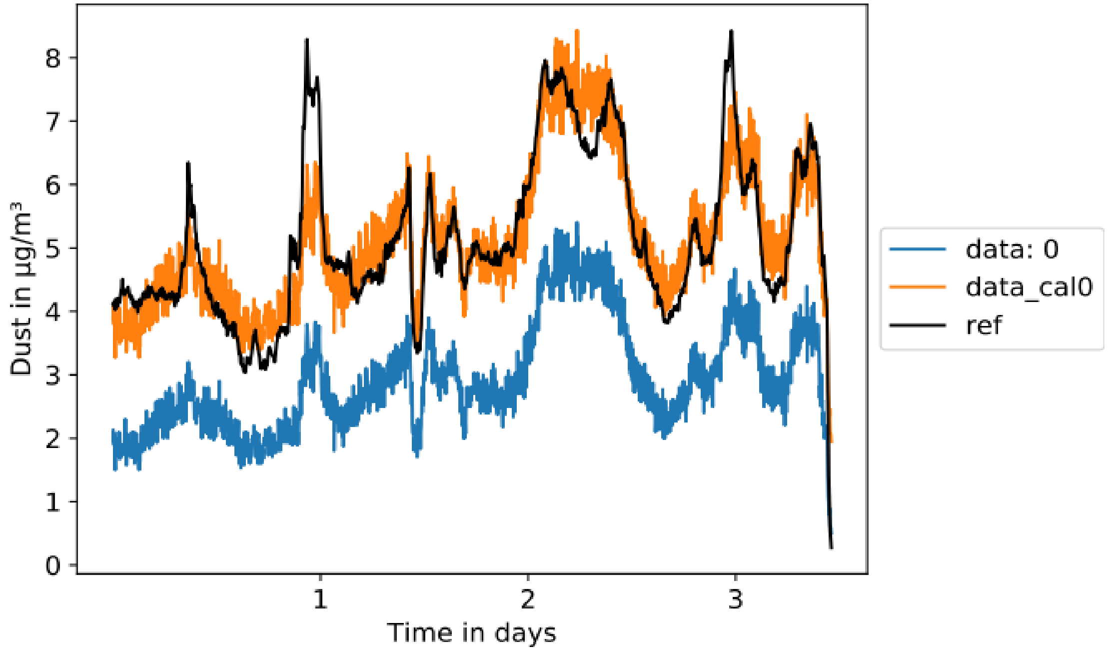


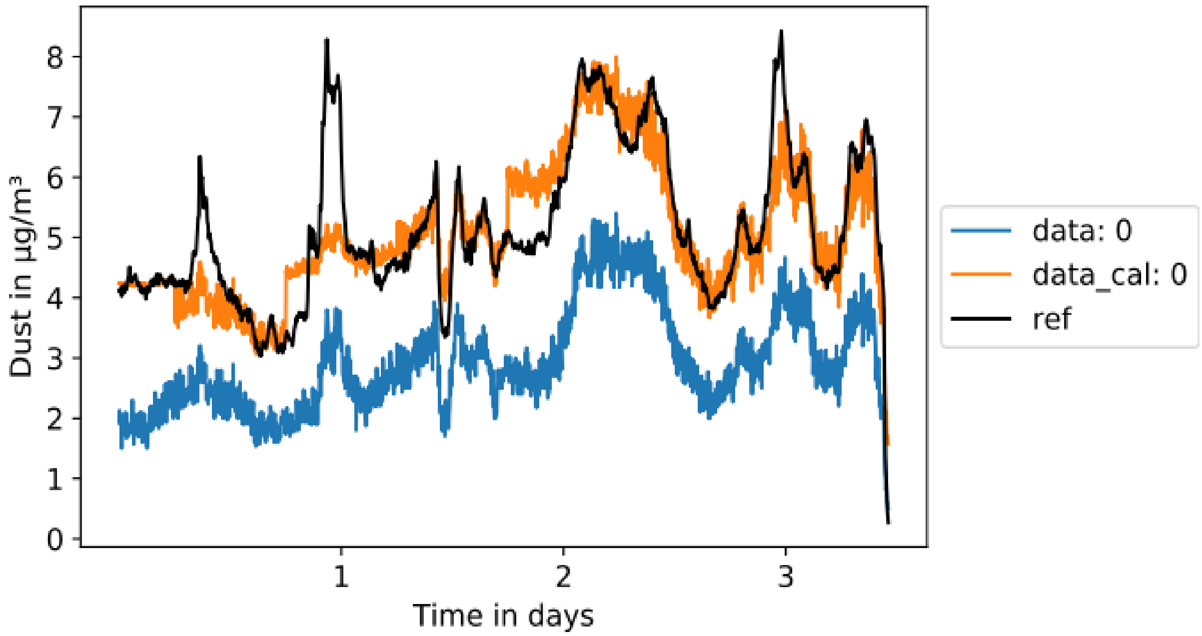
Jose M.Barcelo-Ordinas et al. (2019) Self-calibration methods for uncontrolled environments in sensor networks: A reference survey, Ad Hoc Networks,Volume 88

Multi-hop and rendezvous calibration



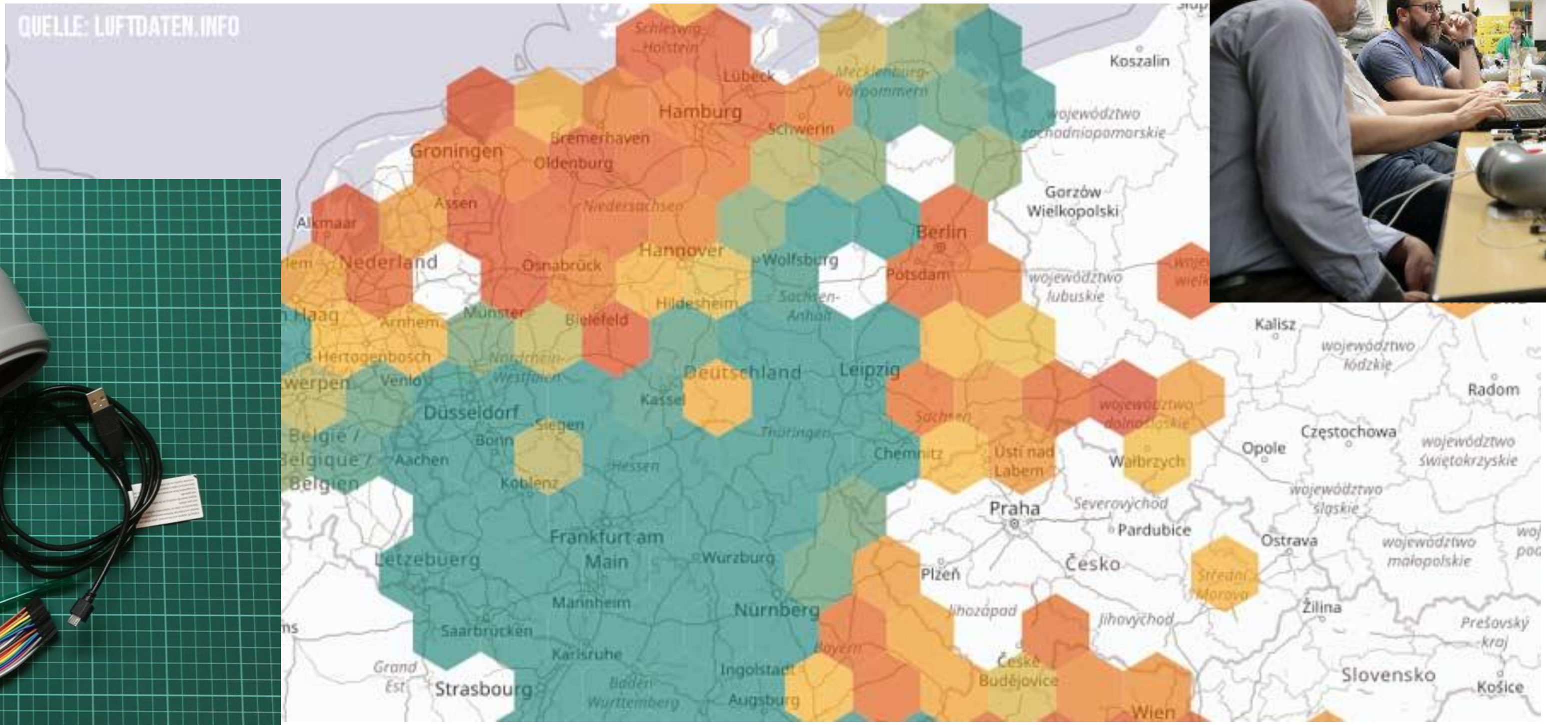
Sometimes less may be more





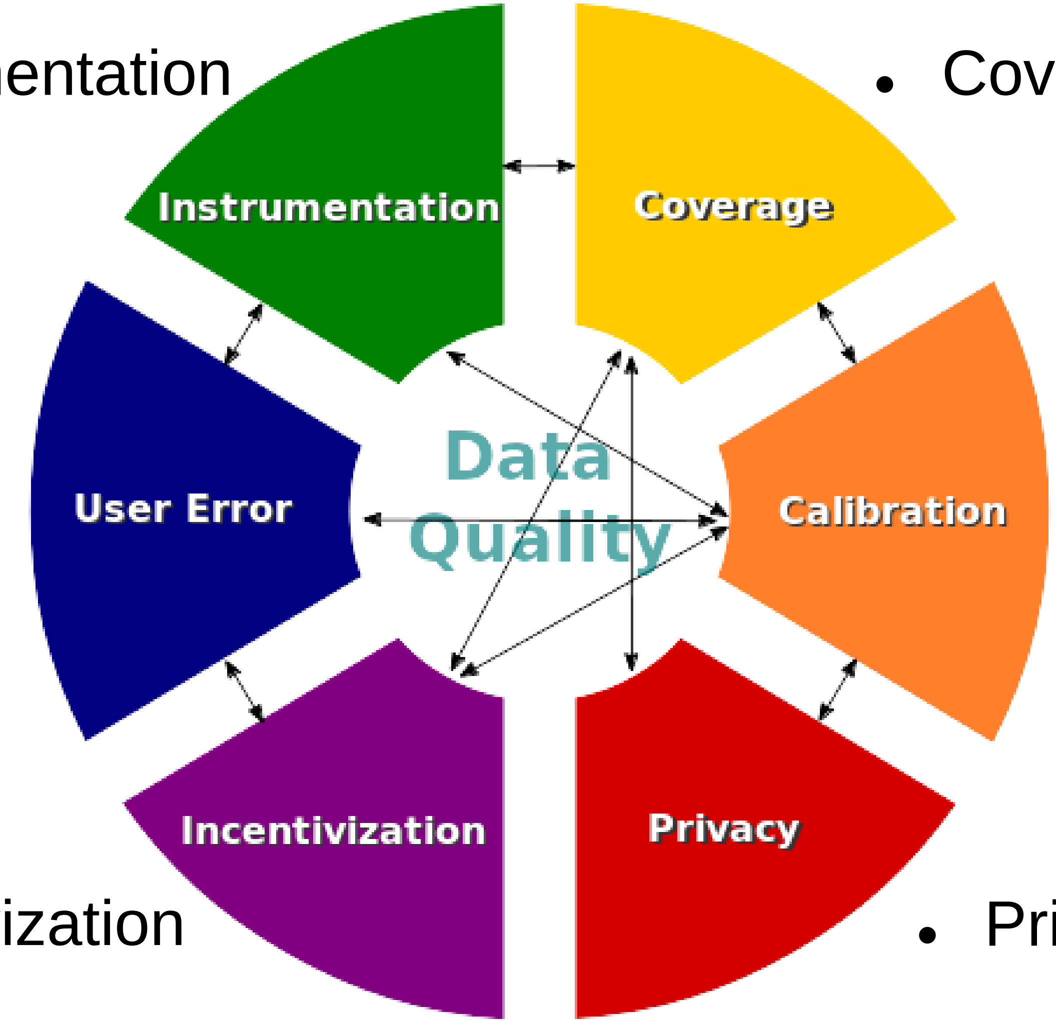
Alternative Deployment?: Citizen Science?

- Sensors assembled and operated by citizens
- Data collected and displayed on luftdaten.info
- Stuttgart alone: more than 600 active sensors



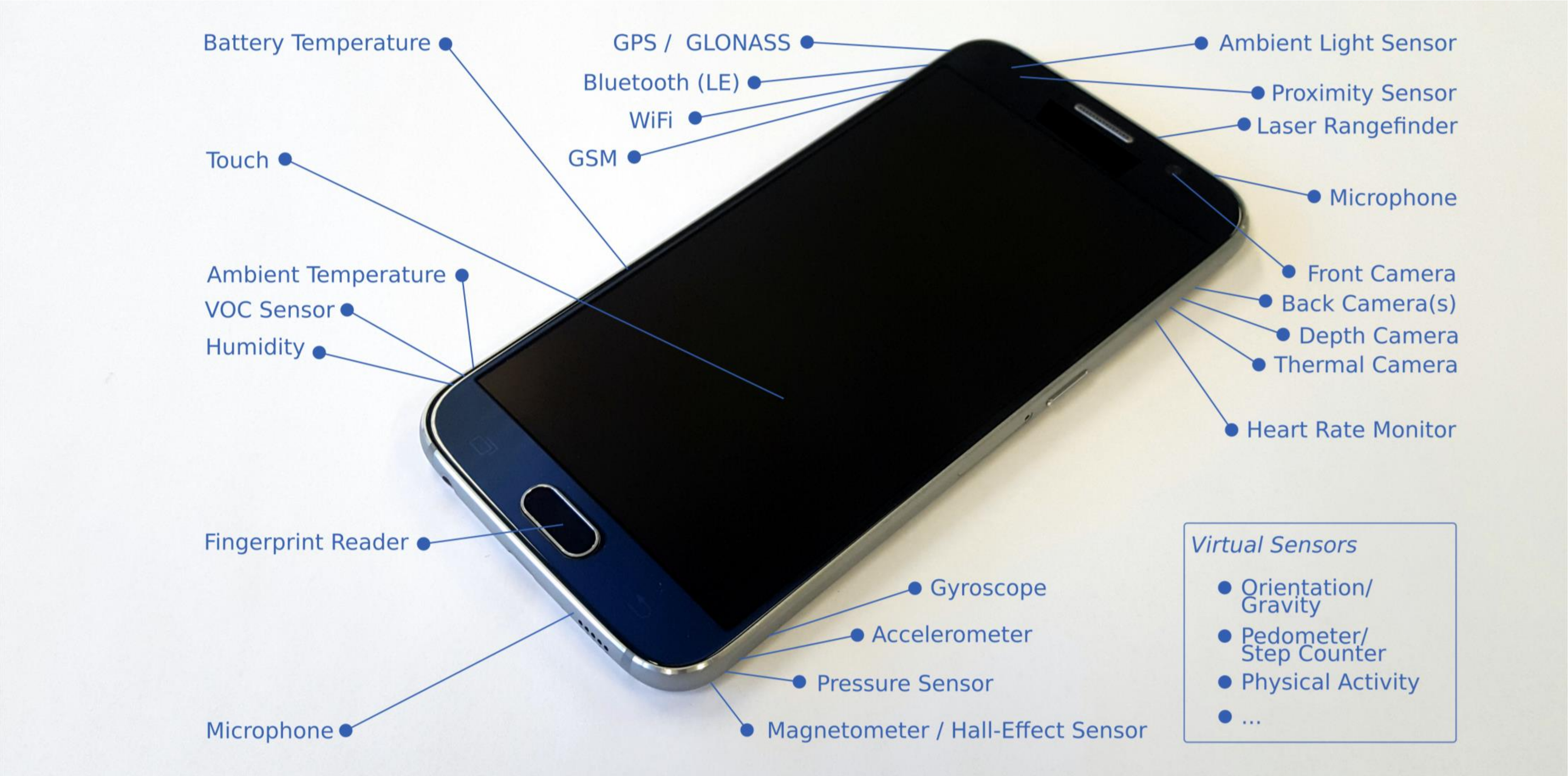
Challenges

- Instrumentation
- Coverage

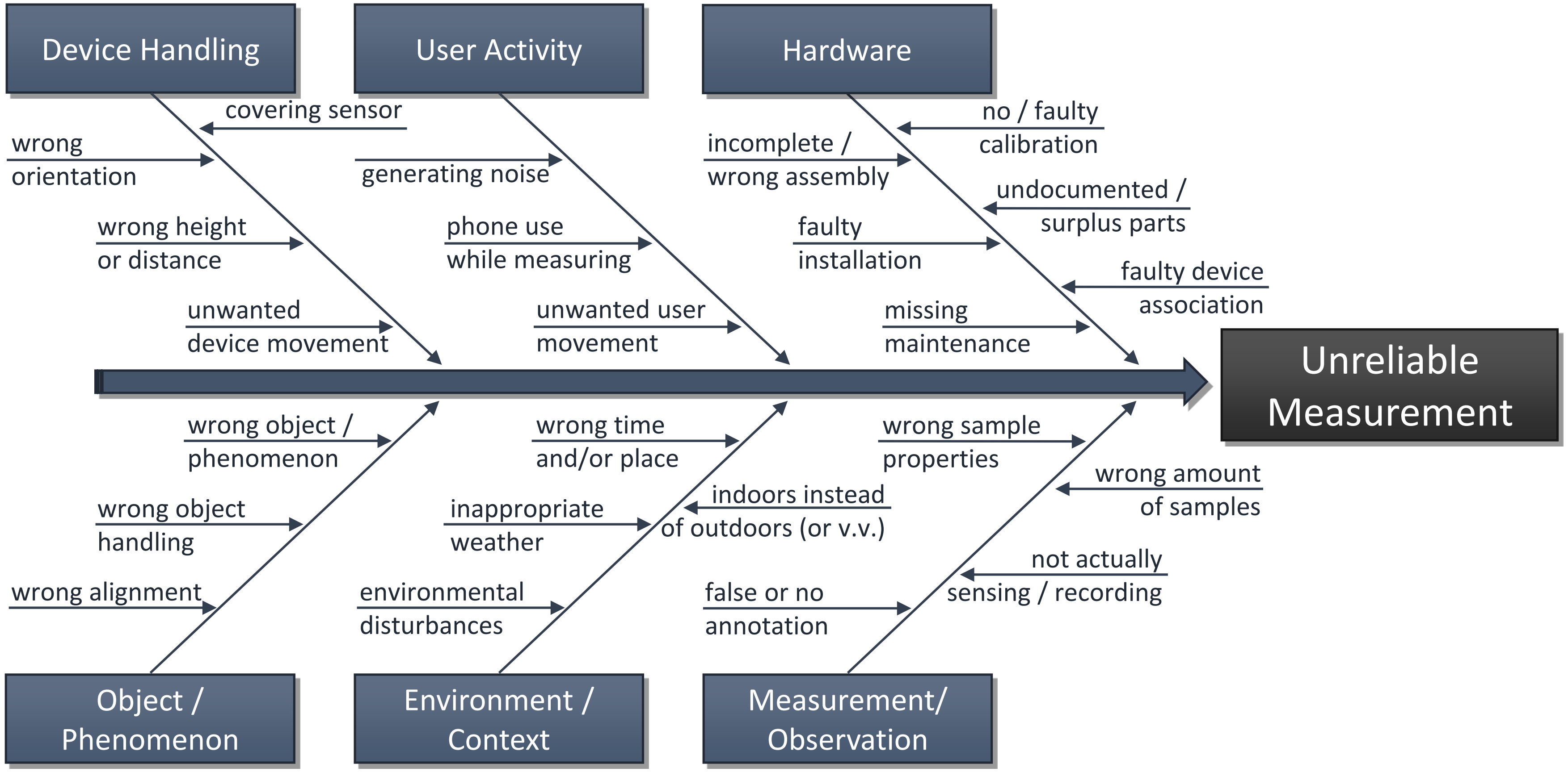


- User
- Error
- Calibration
- Incentivization
- Privacy

Smartphone Sensing

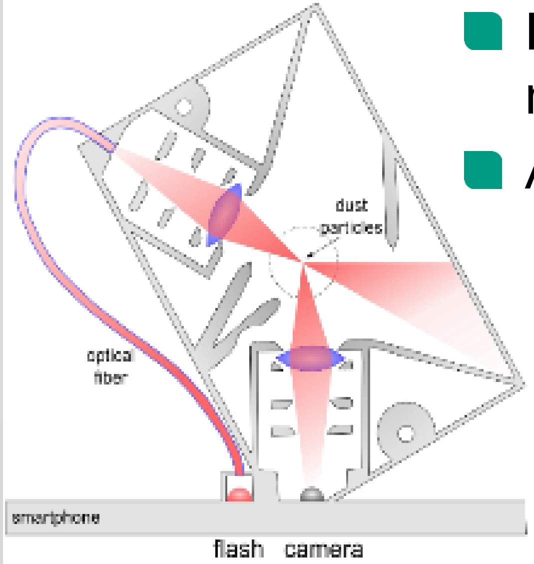


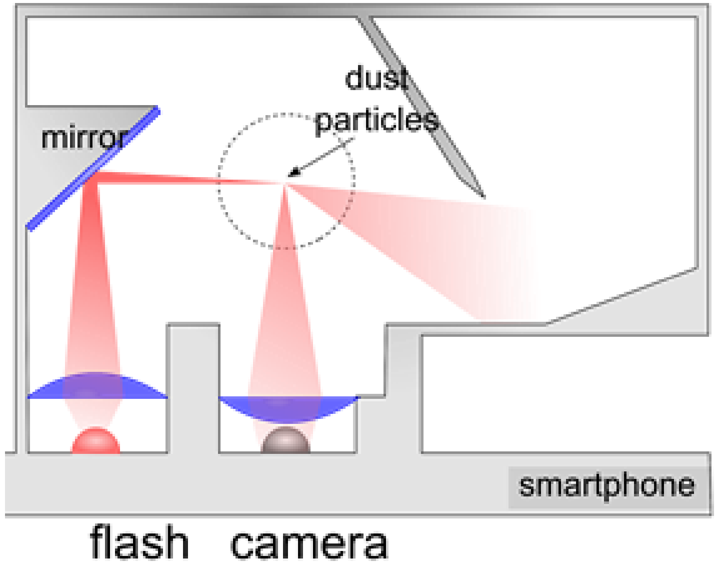
Human Error


Device Handling
User Activity
Hardware
covering sensor
wrong orientation
wrong height or distance
unwanted device movement
generating noise
phone use while measuring
unwanted user movement
no / faulty calibration
incomplete / wrong assembly
undocumented / surplus parts
faulty installation
faulty device association
missing maintenance
Unreliable Measurement
wrong object / phenomenon
wrong object handling
wrong alignment
wrong time and/or place
indoors instead of outdoors (or v.v.)
inappropriate weather
environmental disturbances
wrong sample properties
wrong amount of samples
not actually sensing / recording
false or no annotation
Object / Phenomenon
Environment / Context
Measurement/ Observation

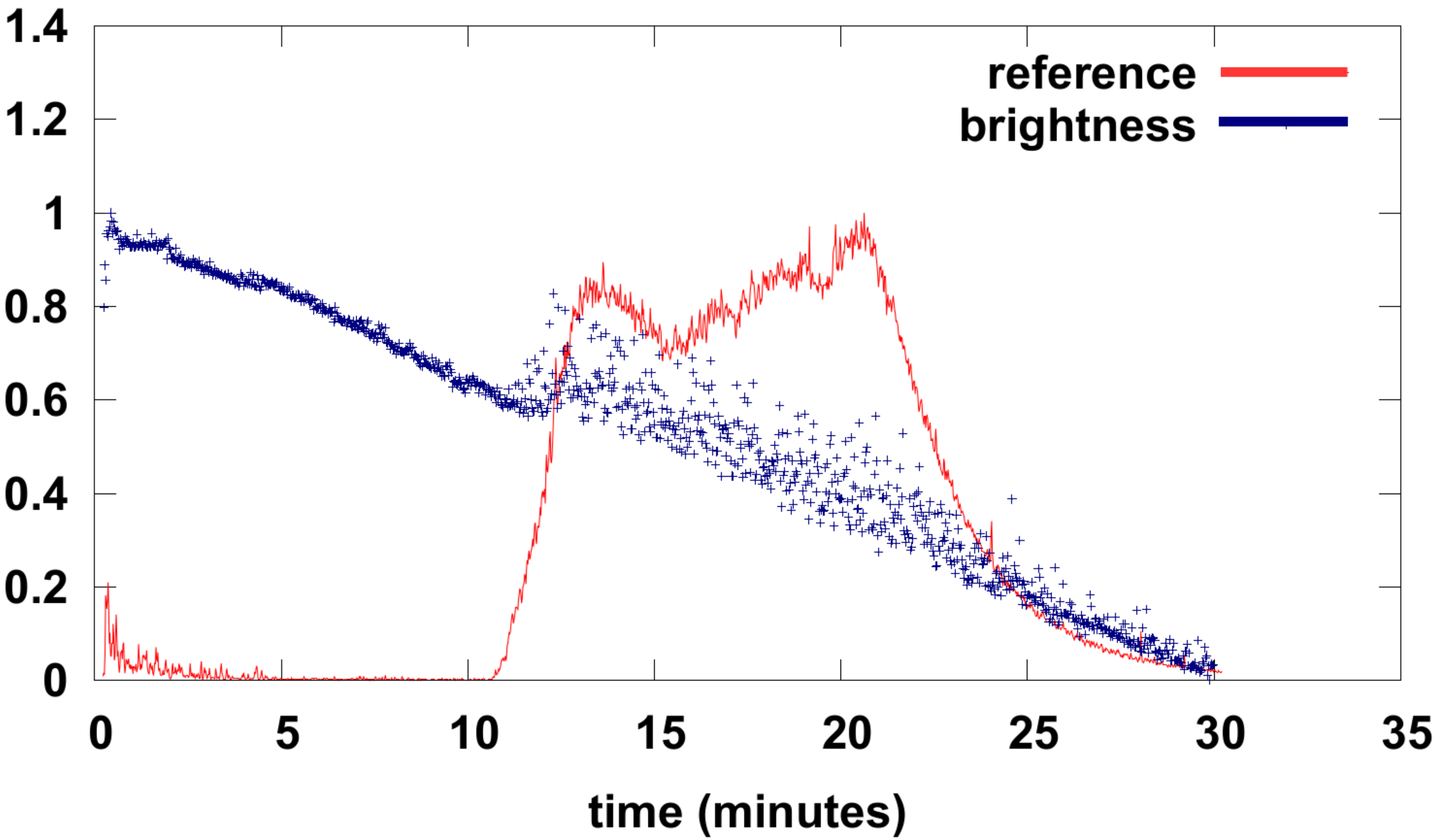
Going cheaper: Smartphone Retrofit PM Sensor

- Idea: Clip-on PM sensor module for smartphones
- 4 generations of prototypes:
  - 3D printed for rapid prototyping
  - Light from flash is rerouted using an optical fiber respectively a mirror
  - Active versions with externally powered LEDs

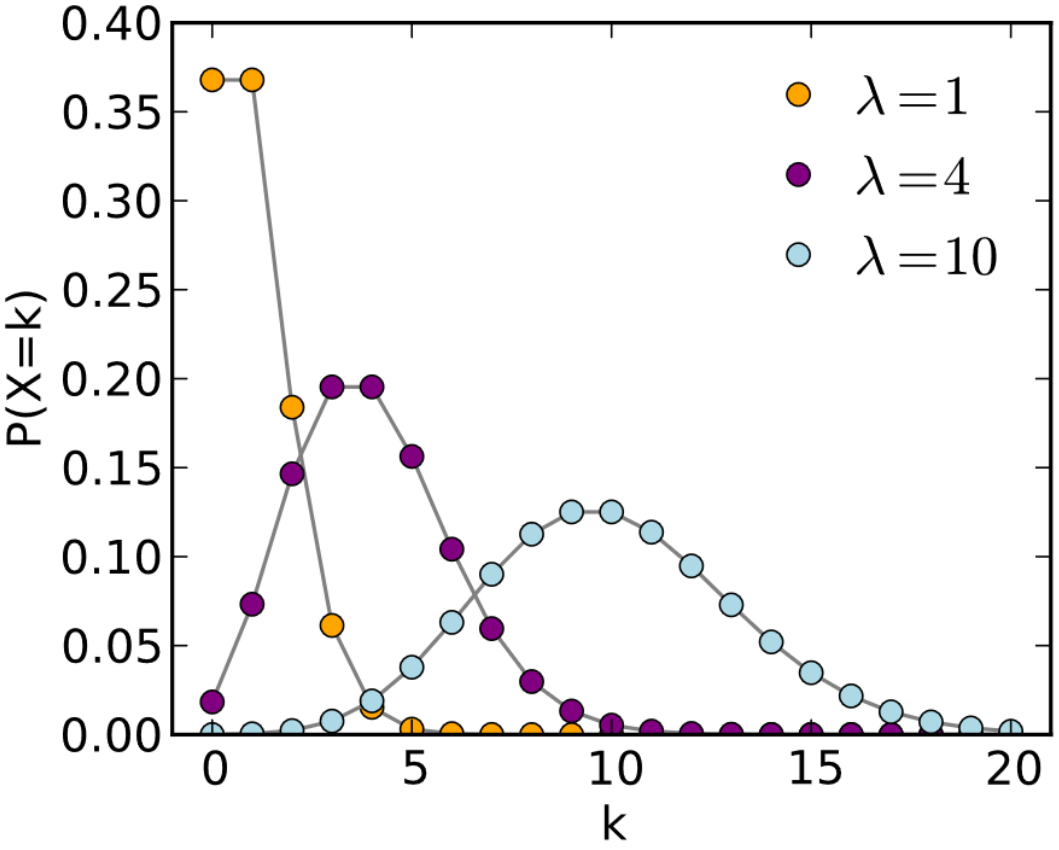




Sensing: Signal Reconstruction from Noise

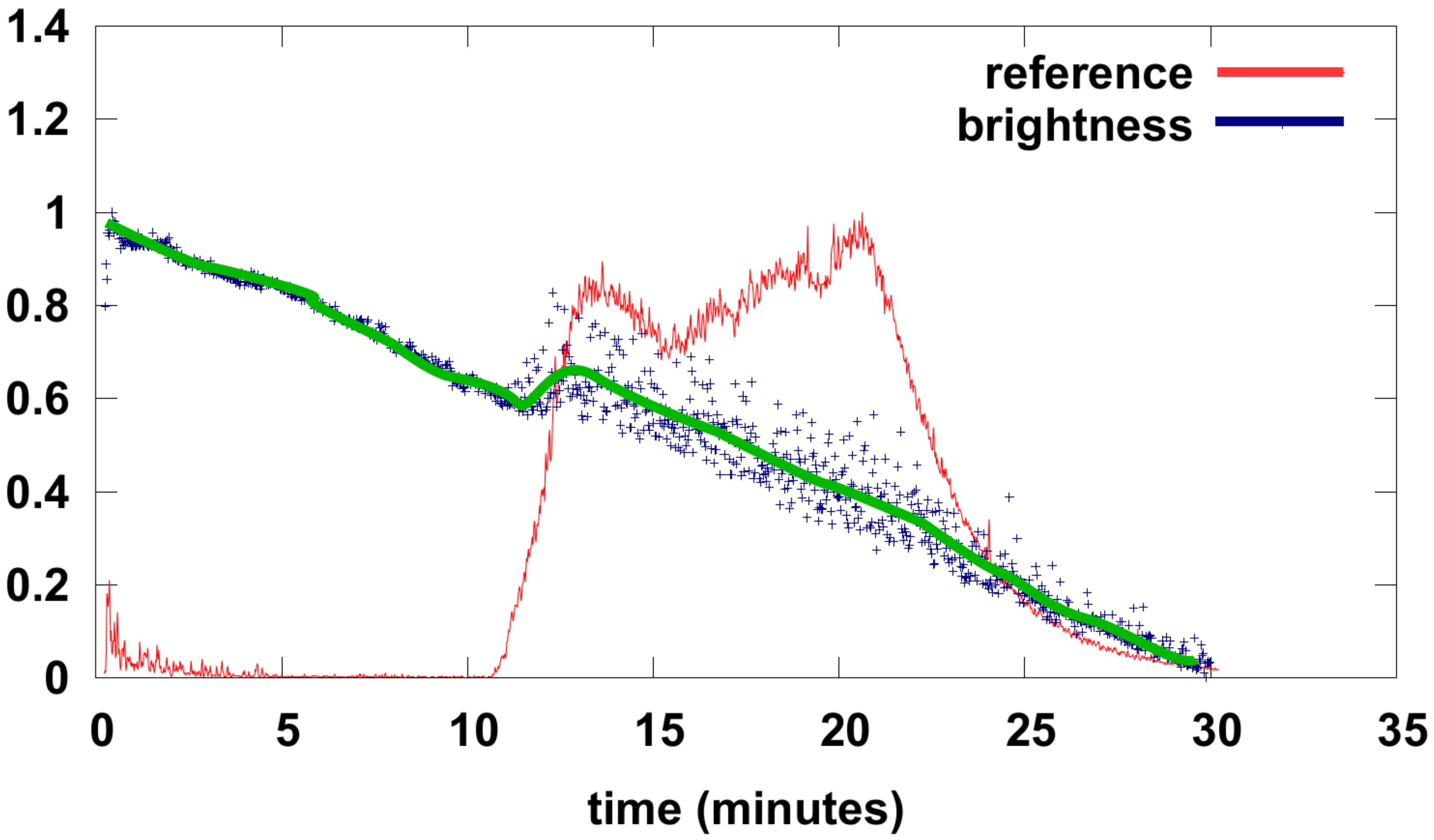


We are *counting* particles…

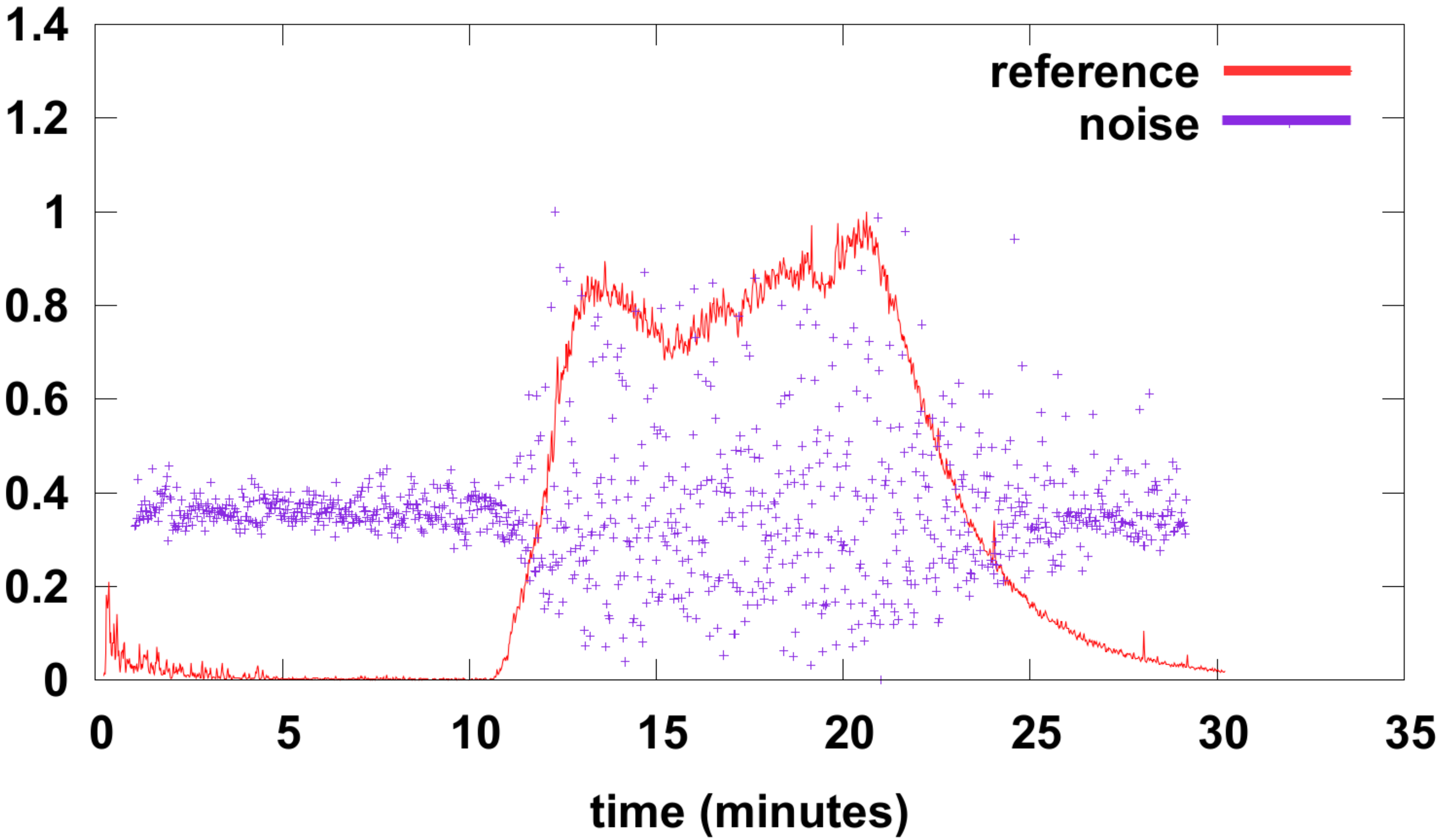


…so we have a relation between mean and variance

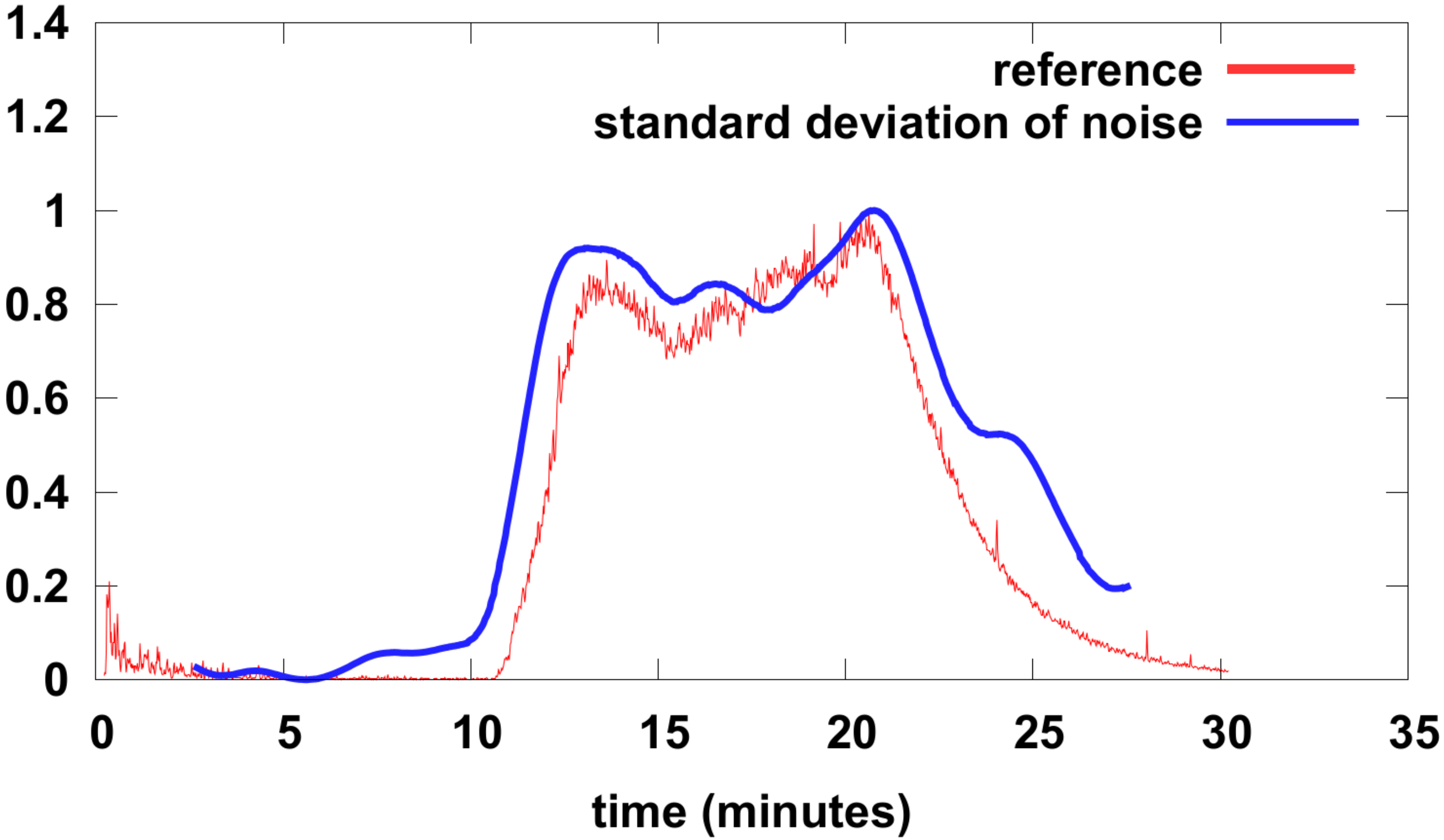
Sensing: Signal Reconstruction from Noise



Sensing: Signal Reconstruction from Noise



Sensing: Signal Reconstruction from Noise



“Low Cost” Learning of PM

Till Riedel, till.riedel@kit.edu

KIT Department of Informatics

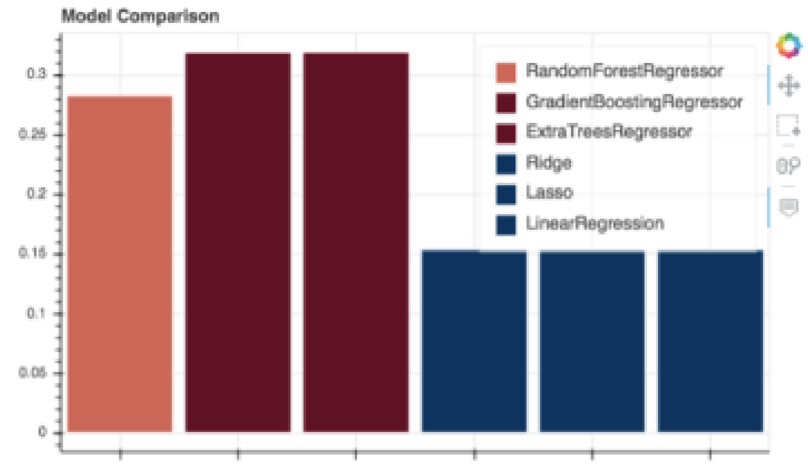
Data analytics applications for Smart Cities combining machine learning methodology, spatial-temporal data analysis and visualization

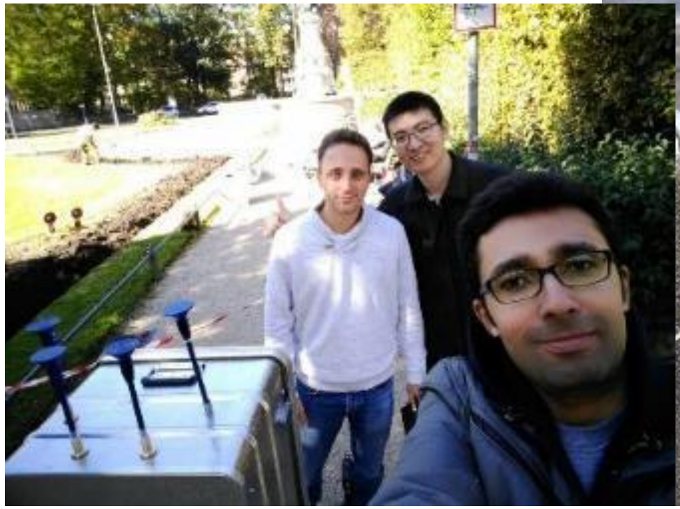
Masterarbeit
von

Yao Shen

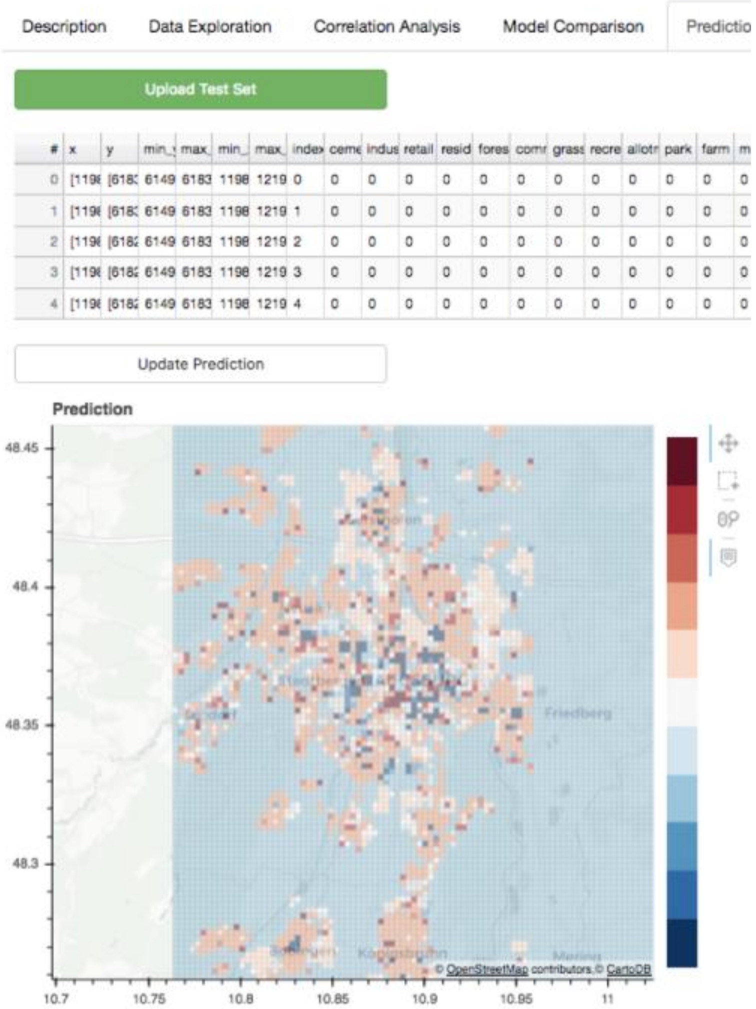
EIFER
EUROPEAN INSTITUTE FOR ENERGY RESEARCH

KIT
Karlsruher Institut für Technologie



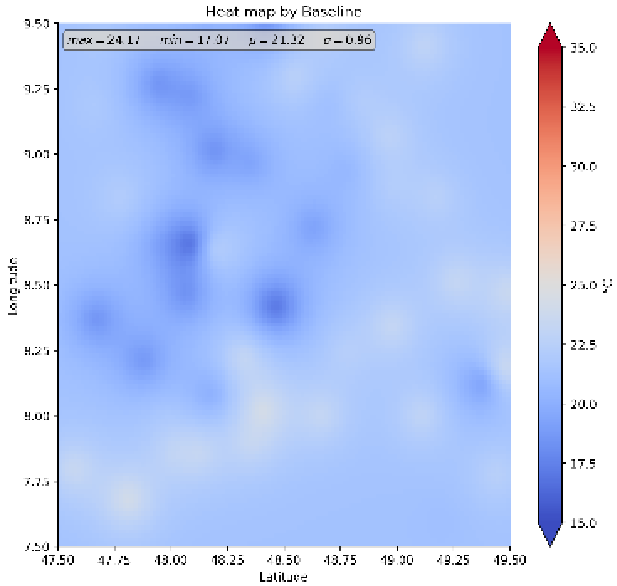


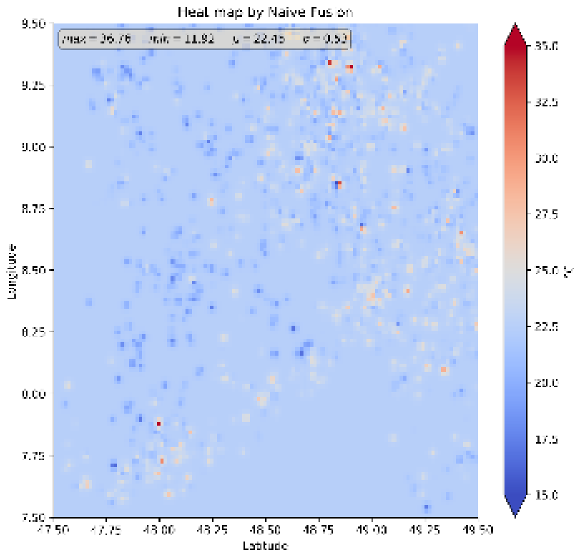
HelmholtzZentrum münchen
German Research Center for Environmental Health

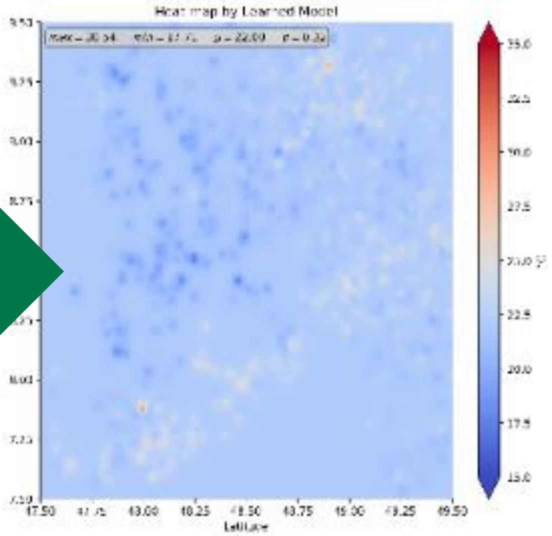


Automated Quality Assessment of (Cititzen) Weather Stations

- **Julian Bruns**, Johannes Riesterer, Bowen Wang, Till Riedel, Michael Beigl, GI_Forum 2018







Weather Stations Baden-Württenberg

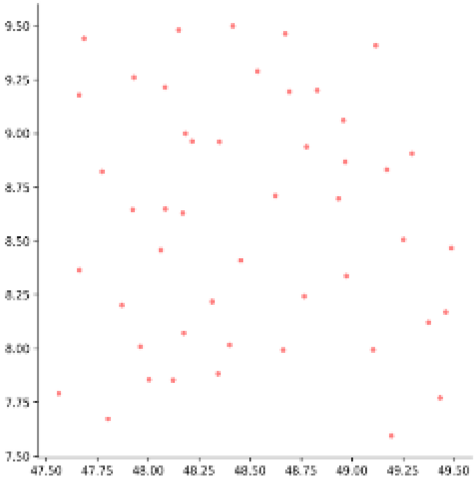
Karlsruher Institut für Technologie

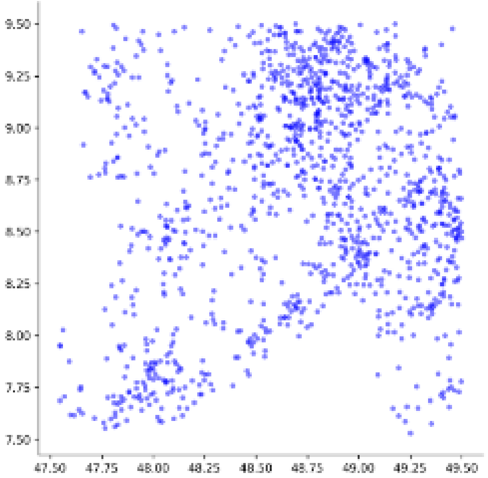
DWD
Deutscher Wetterdienst
Wetter und Klima aus einer Hand

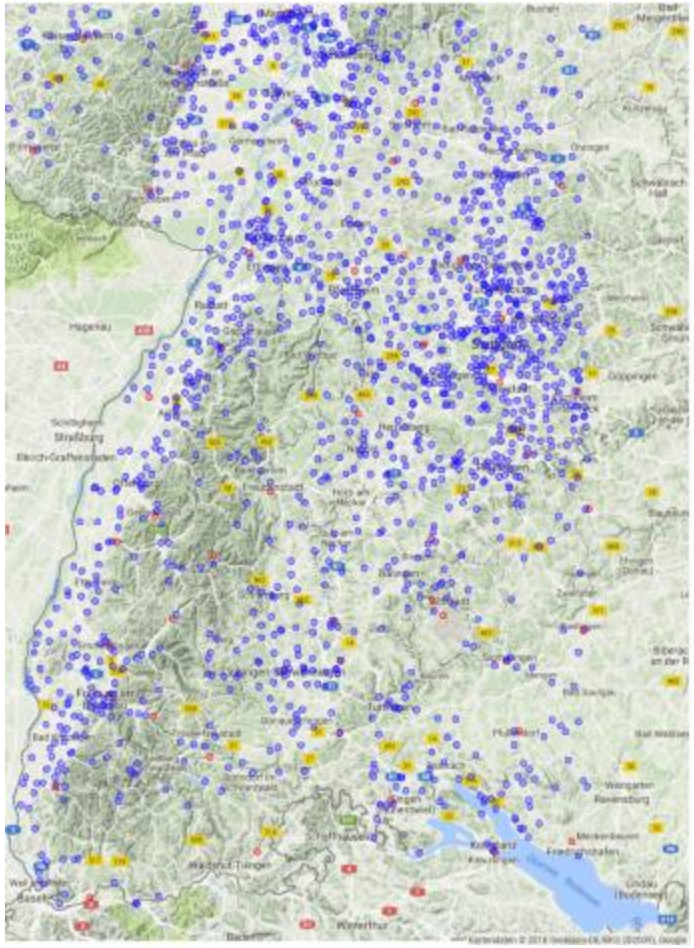
- Authoritative
- High Quality
- Low Density

WEATHER UNDERGROUND

- Crowd-Sourced
- Low Quality
- High Density







Model Background

- Kriging (Krige 1951)
  - A method of interpolation

$$y^*(\vec{x}) = \sum_{i=1}^{n} \omega_i \times y_i \tag{4.10}$$

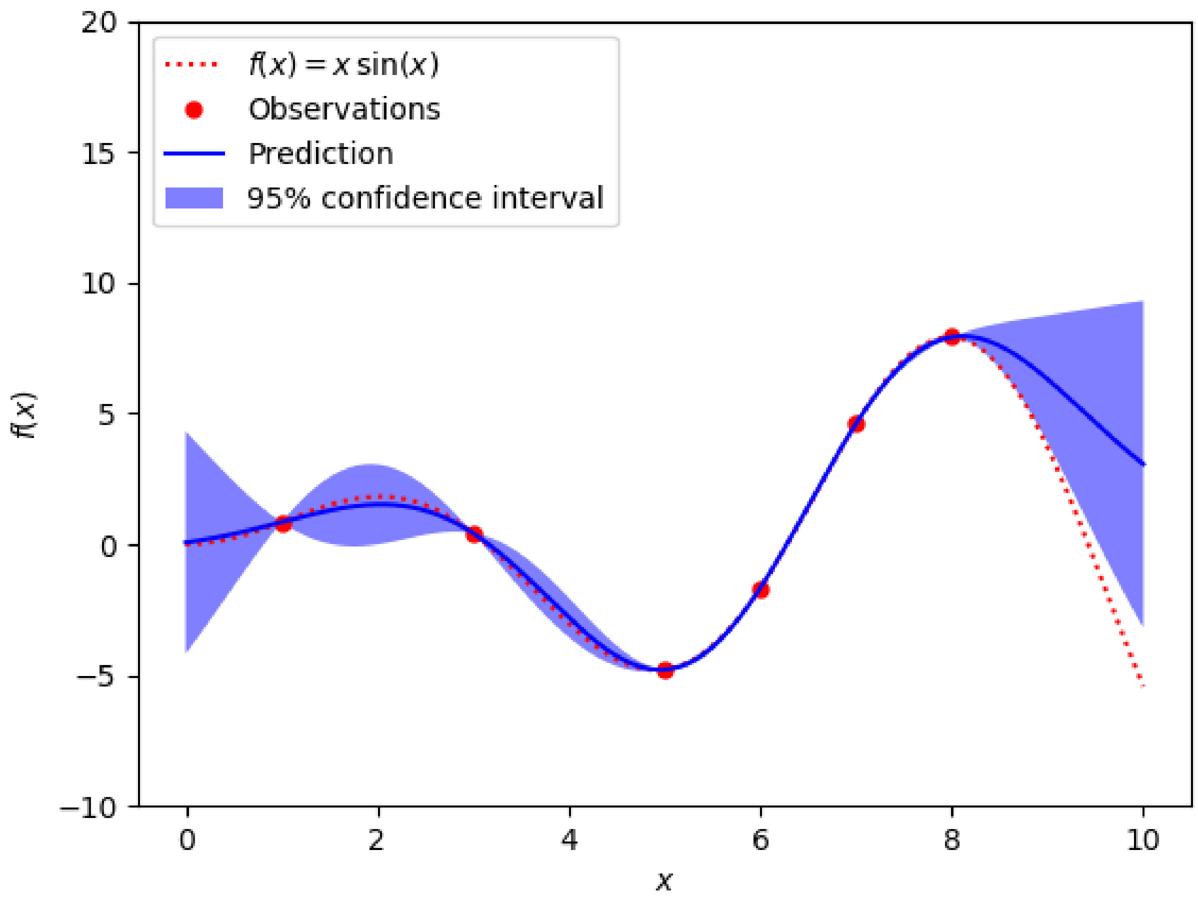
  - Correlation depends on “distance”

$$K = \begin{bmatrix} k_{x_0,x_0} & k_{x_1,x_0} & \dots & k_{x_{n-1},x_0} \\ k_{x_0,x_1} & k_{x_1,x_1} & \dots & k_{x_{n-1},x_1} \\ \vdots & k_{x_i,x_j} & \ddots & \vdots \\ k_{x_0,x_{n-1}} & \dots & k_{x_{n-2},x_{n-1}} & k_{x_{n-1},x_{n-1}} \end{bmatrix} \tag{4.11}$$

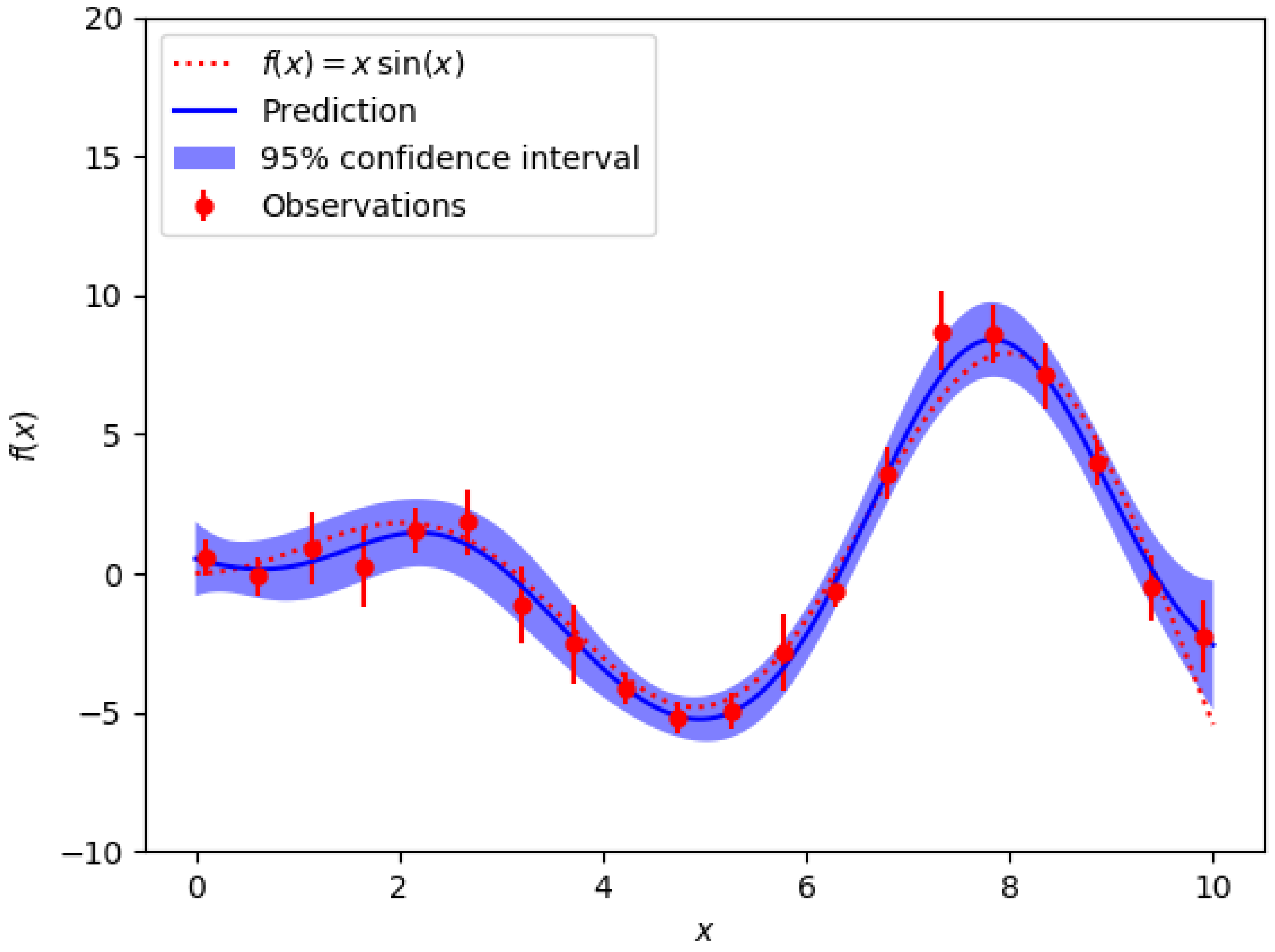
$$K_* = \begin{bmatrix} k_{x_*,x_0} & k_{x_*,x_1} & \dots & k_{x_*,x_{n-1}} \end{bmatrix} \tag{4.12}$$

$$K_{**} = k(x_*, x_*) \tag{4.13}$$

  - “Distance” can defined by kernel function → Here we could include physics!!
  - → Leads to **Gaussian Process Regression** (Edward et al. 2006)


$f(x) = x\sin(x)$
Observations
Prediction
95% confidence interval
20
15
10
5
0
−5
−10
0
2
4
6
8
10
x
$f(x)$

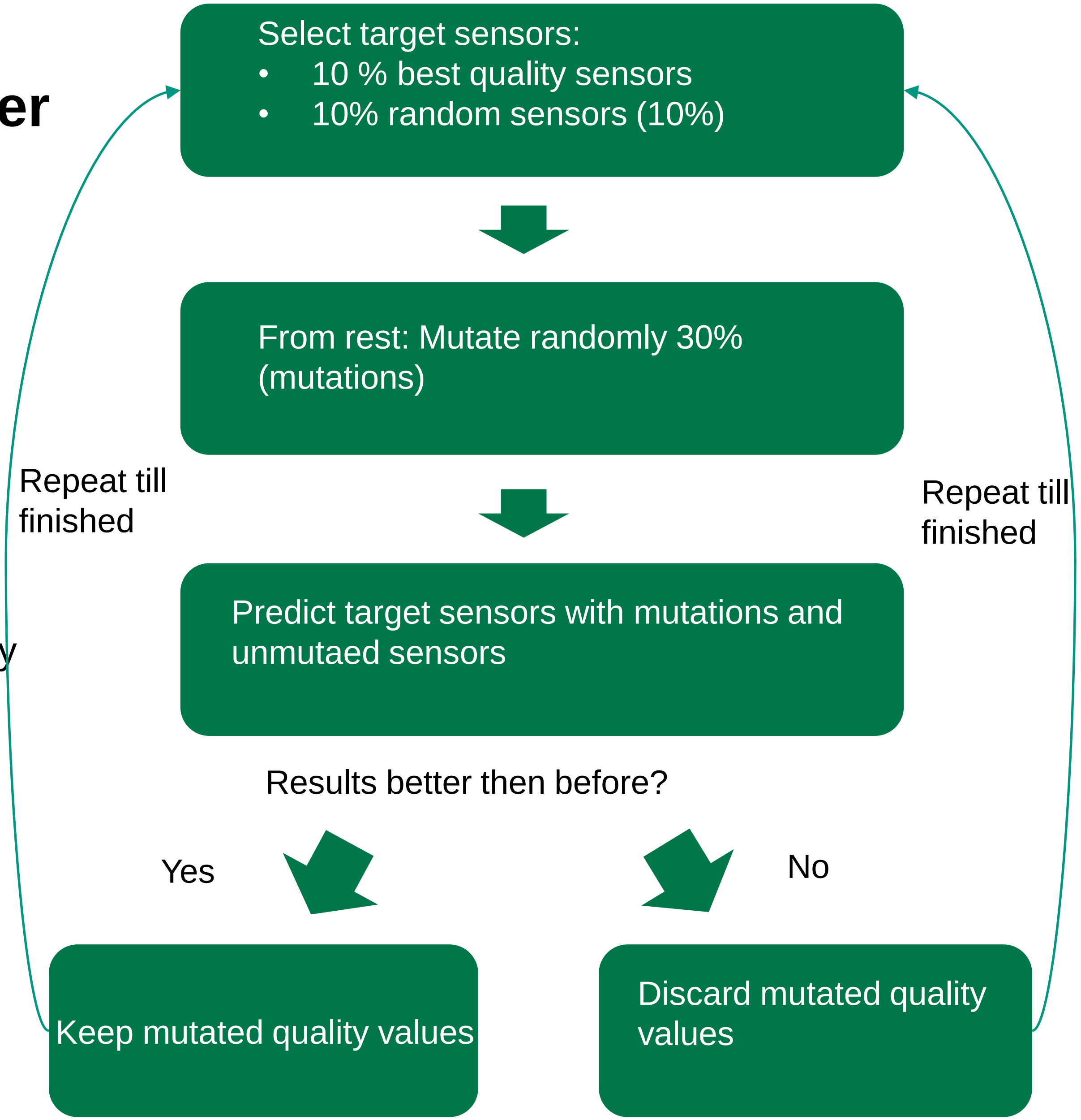
Biggest problem: how to assess confidence of measurement



This is how computer science does it: What doesn't fit is made to fit

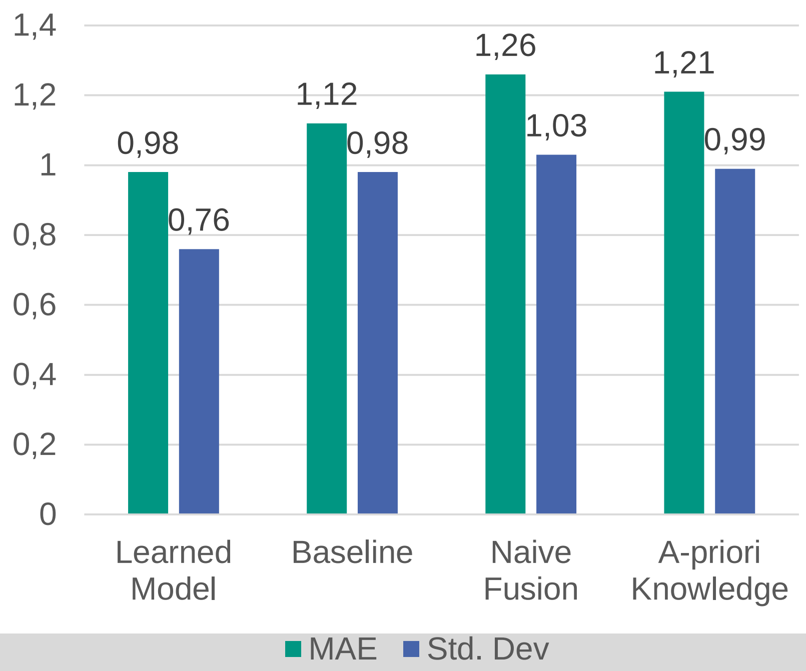
Genetic Algorithm (Black-box-optimization)

- How to learn the individual Quality?
- Randomly mutated the quality values by 0.1
- Learn GPR
- Compare results to results before last iteration
- Keep better result
- Repeat



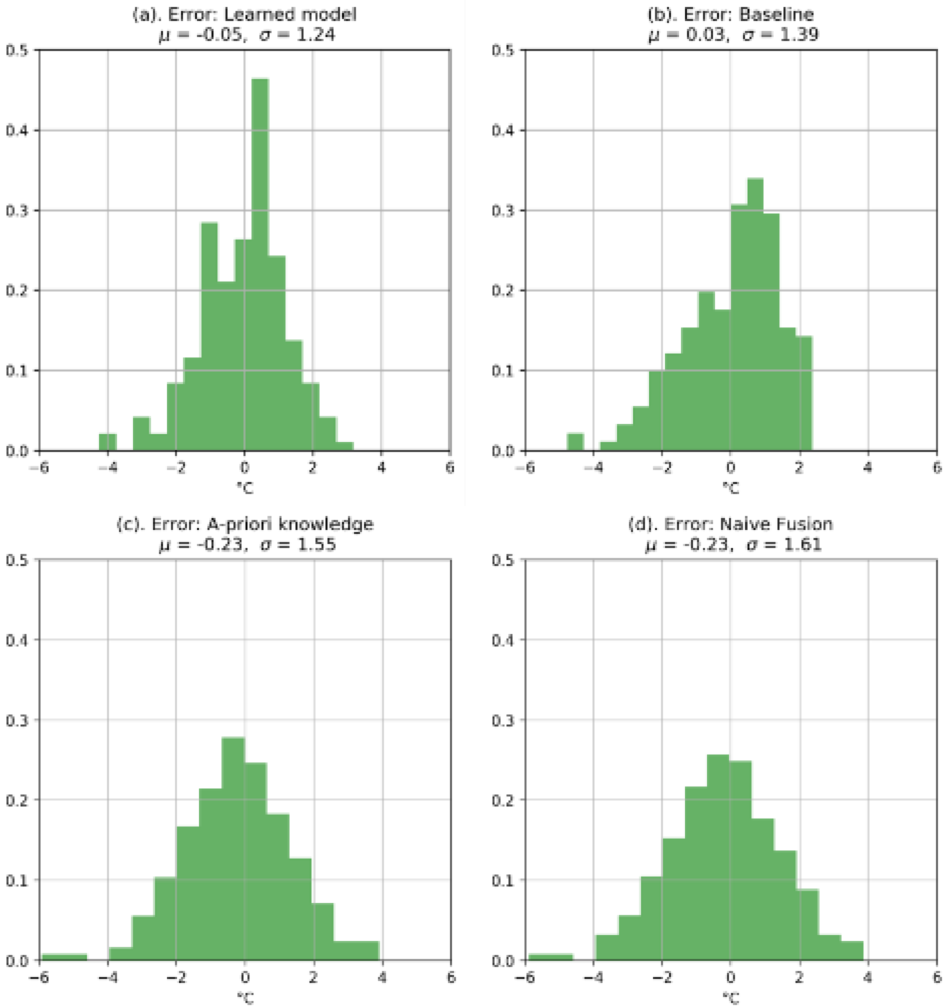
Accuracy Test

- Task: predicting the air temperature from DWD weather stations
- Data set:
  - The observations in August 2016 at 12 O’Clock
  - 42,966 observations generated by 1,561 weather stations



MODEL	MEAN ABSOLUTE ERROR	STANDARD DEVIATION
BASELINE	1.12°K	0.83°K
NAÏVE FUSION	1.26°K (-12.5%)	1.03°K
A-PRIORI INFORMATION	1.21°K (-8.0%)	0.99°K
LEARNED MODEL	0.98°K (12.5%)	0.76°K

Histogram of the Prediction Error Out Of Sample



- Error of VGI is ~ normal distributed
- DWD stations induce smaller error, but Bias in form of positive error
- Learned Model strongly reduces variance and MAE
- Highest errors are negative
  - Indicator for reference station with relatively low temperature in regard to surrounding area

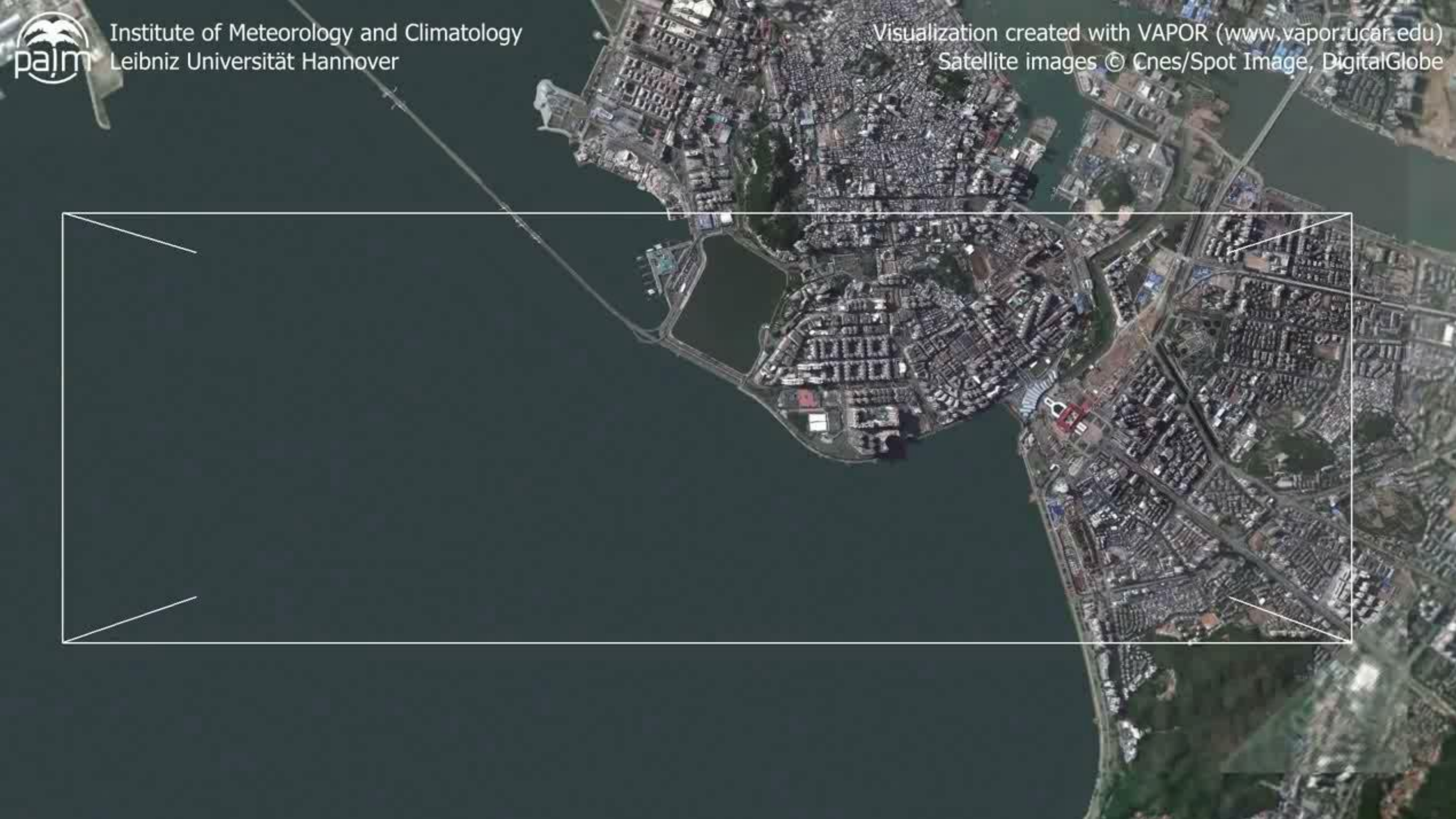
This when computer scientists do things

- Computer scientists don‘t care as long it works and its fast (parallelizable)
- Would be much better to model optimization of uncertainty inside gaussian process…
- However, at what price are we introducing additional complexity?

Limits of predictive analysis (ML-based AI)

- We can only extrapolate from the past to build a looking glas into the future → work if the system stays stable
- We cannot tell what happens if the input data is from a different context → ML models tend to overfit
- ML models are difficult to „fix“ or to deeply understand → we understand the learning algorithm but not always the path it takes

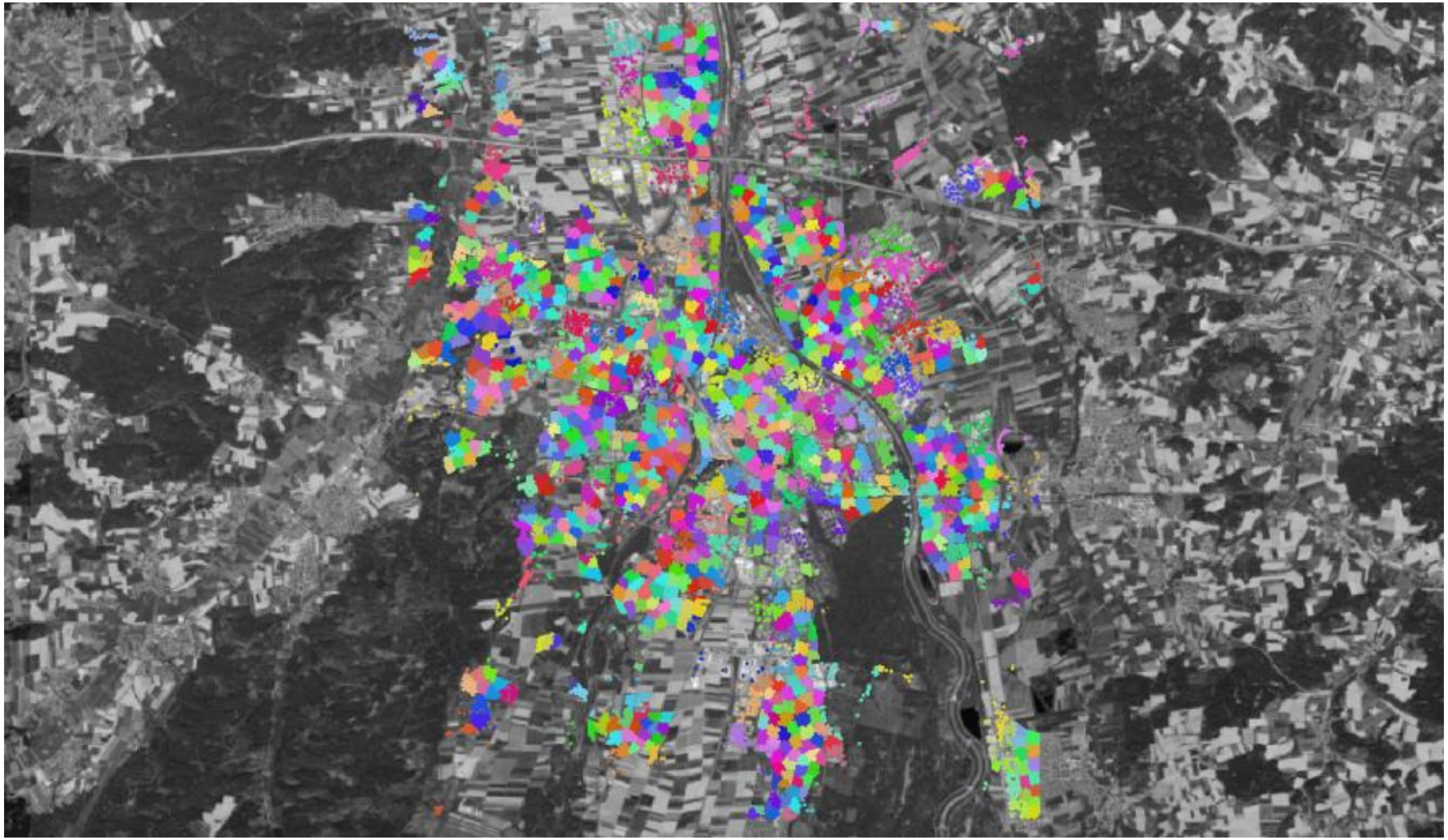
➔ Only prescriptive model


palm
Institute of Meteorology and Climatology
Leibniz Universität Hannover
Visualization created with VAPOR (www.vapor.ucar.edu)
Satellite images © Cnes/Spot Image, DigitalGlobe

But also Modelling needs measurement!

- Emission inventory: Point sources
  - Industrial emissions not transparent
  - Working together with chimney cleaners
    - Privacy issues set limits on spatial data granularity
  - Generating time granularity (modulation) is „educated guessing“
  - Very expensive because extra effort/no standardization
- Emision inventory: Traffic emission
  - Floating Car Traffic loop data (not calibrated)
  - Only major roads covered, traffic models often too course granular
- Metereological information
  - Few measurement station to initialize
  - Wind Lidars/ Ceilometers / RAS not constantly operated / available
- 3D-Models of the city
  - Land use and LOD2 Data
  - Remote sensing data, open streetmaps are alternatives

Clustering of emission sources in Augsburg to maintain privacy

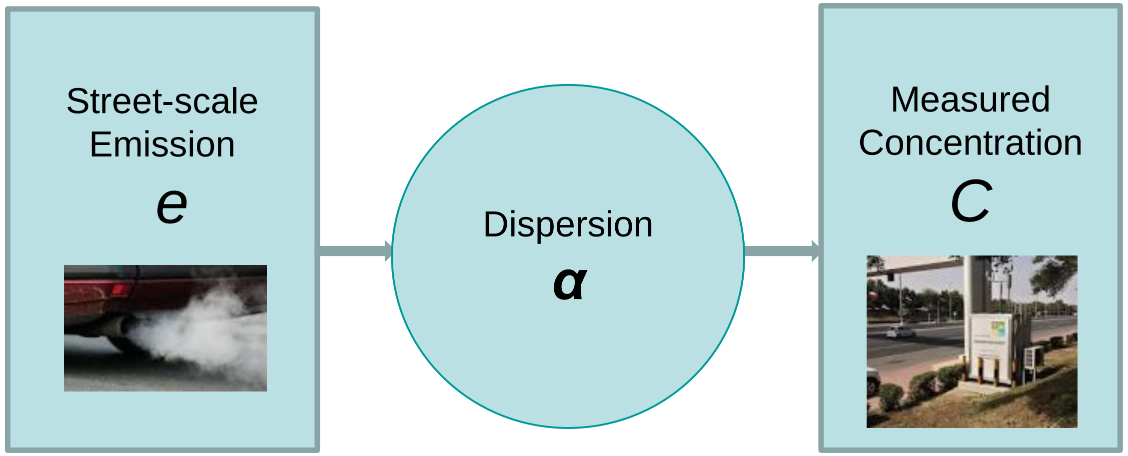


Not either or (some outlook)

- Model assimilation
  - Black-box optimization of hyperparameters
- Inverse modelling
  - Physically informed ML
- Learning physical models
  - Similar boundary condition should lead to similar results

- Superresolution
  - „Smart“ interpolation/debluring in both time and space
- Nowcasting / Extrapolation

Emissions estimation using inverse dispersion modelling



$C_i = a_i e_i$

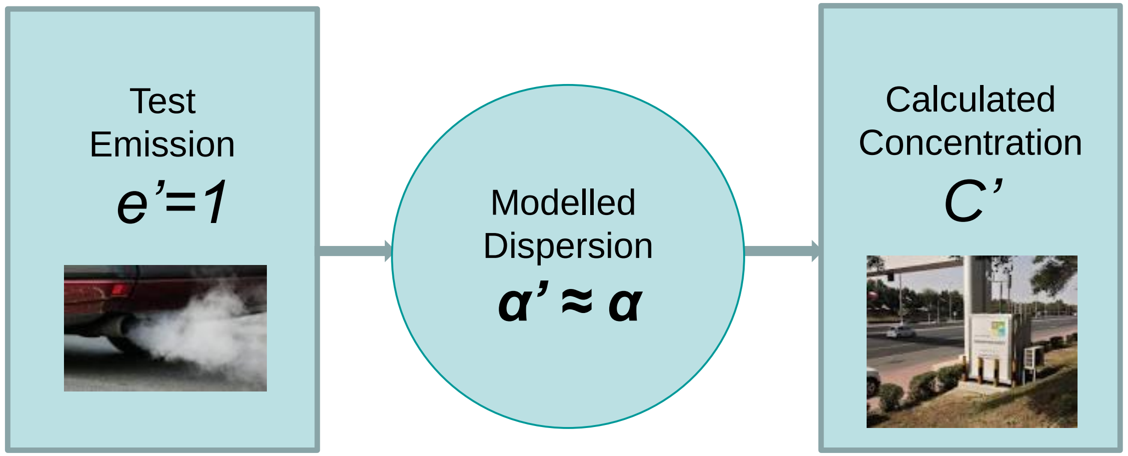
$\boldsymbol{a_i}$ **= dispersion factor of source** $\boldsymbol{i}$

$\boldsymbol{e_i}$ **= emission rate g/s**

C_i = known,

a_i =? e_i = ?

Emissions estimation using inverse dispersion modelling



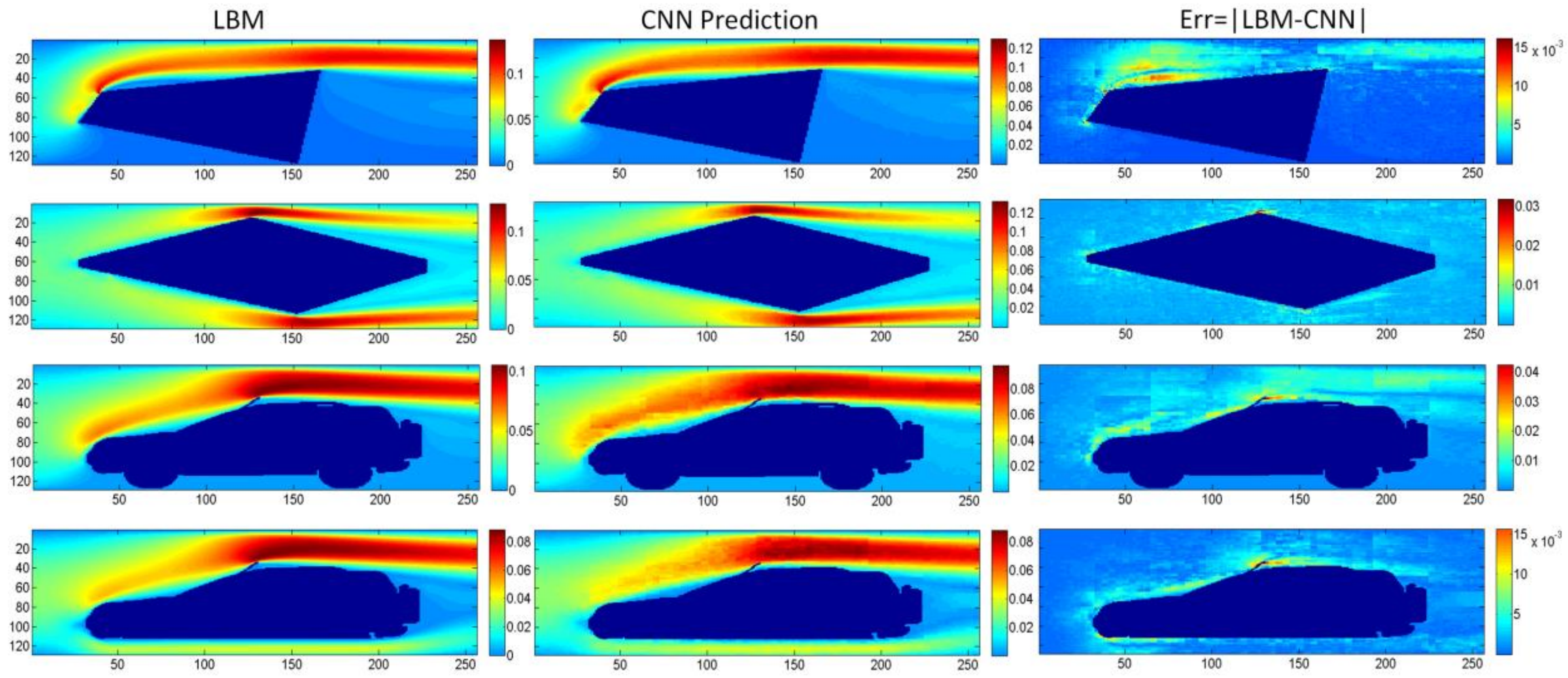
$$e = \frac{C - C_{bg}}{C' - C_{bg}} \cdot e'$$

Convolutional Neural Networks for Steady Flow Approximation

Xiaoxiao Guo
University of Michigan
guoxiao@umich.edu

Wei Li
Autodesk Research
Wei.Li@autodesk.com

Francesco Iorio
Autodesk Research
Francesco.Iorio@autodesk.com



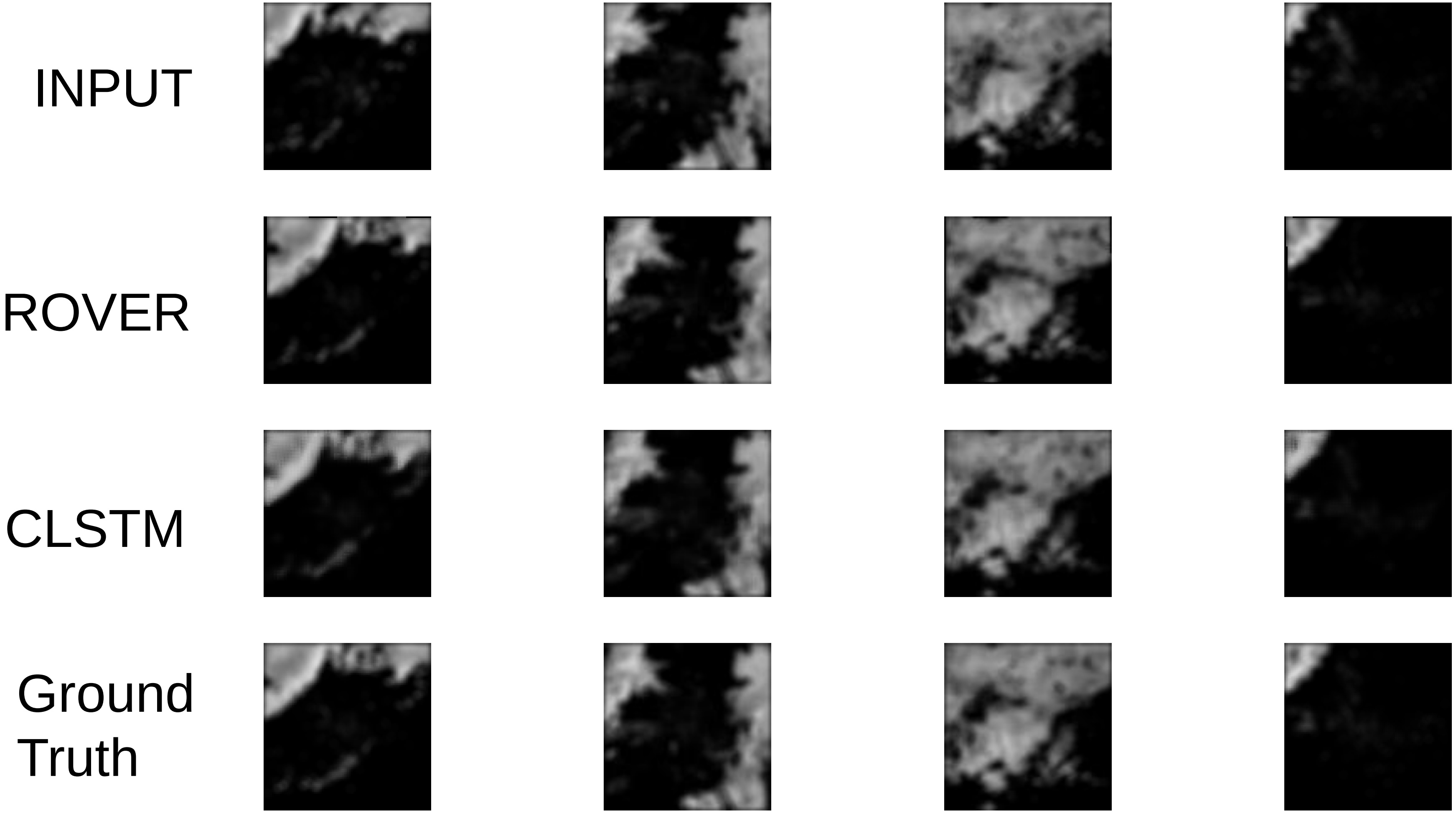
KDD '16, August 13-17, 2016, San Francisco, CA, USA

ISBN 978-1-4503-4232-2/16/08...$15.00
DOI: http://dx.doi.org/10.1145/2939672.2939738

Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting

Xingjian Shi Zhourong Chen Hao Wang Dit-Yan Yeung
Department of Computer Science and Engineering
Hong Kong University of Science and Technology
{xshiab,zchenbb,hwangaz,dyyeung}@cse.ust.hk

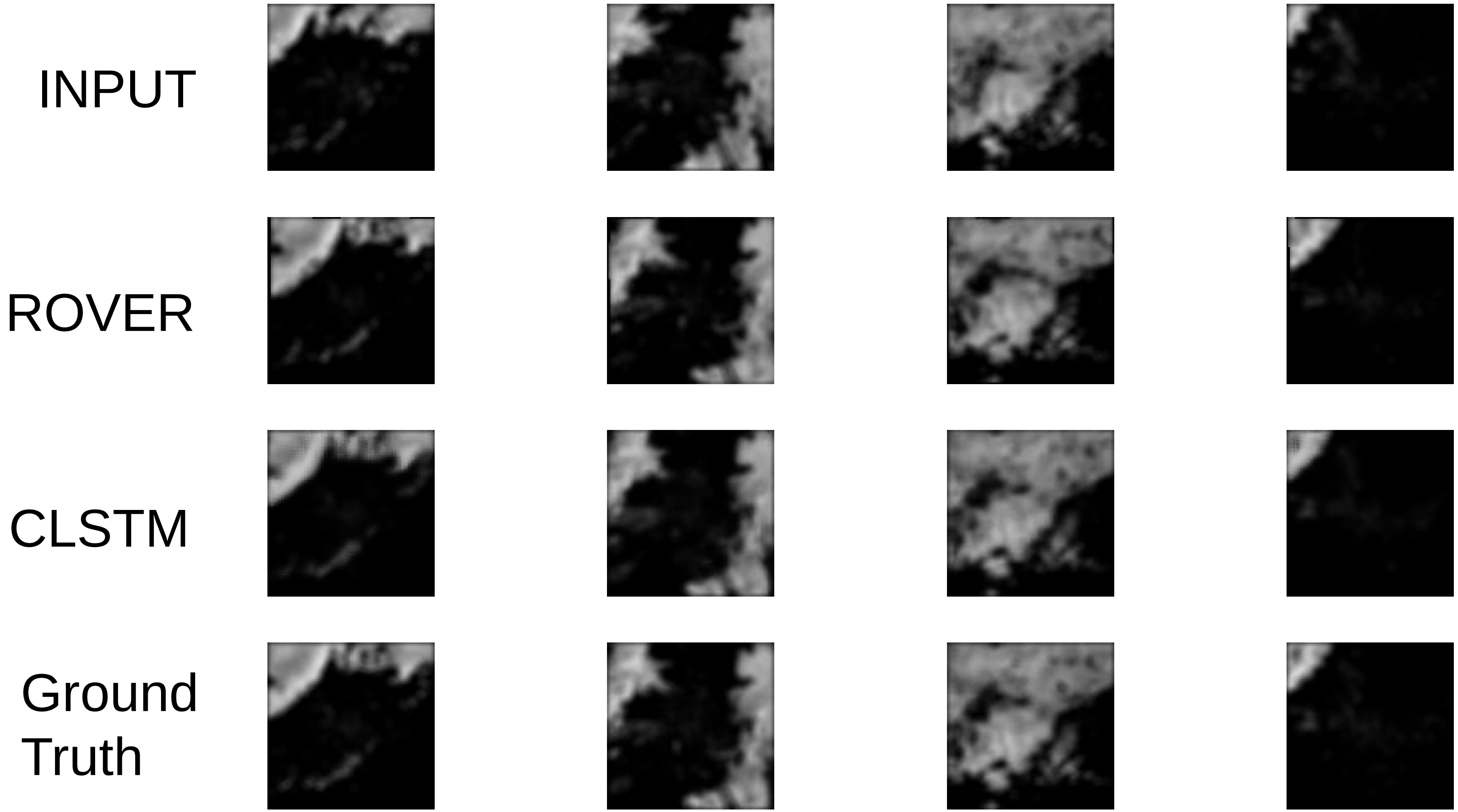
Wai-kin Wong Wang-chun Woo
Hong Kong Observatory
Hong Kong, China
{wkwong,wcwoo}@hko.gov.hk



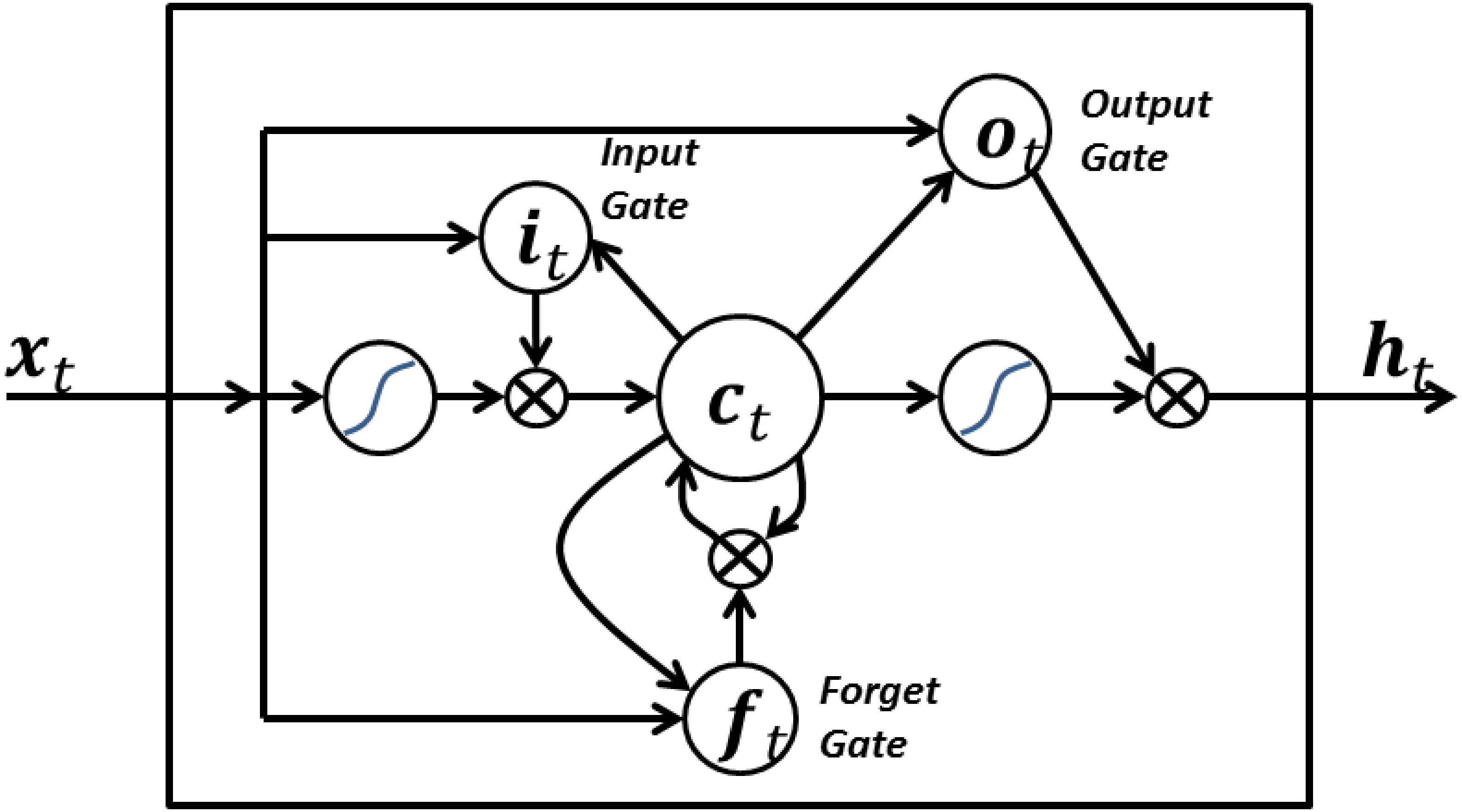
Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting

Xingjian Shi Zhourong Chen Hao Wang Dit-Yan Yeung
Department of Computer Science and Engineering
Hong Kong University of Science and Technology
{xshiab,zchenbb,hwangaz,dyyeung}@cse.ust.hk

Wai-kin Wong Wang-chun Woo
Hong Kong Observatory
Hong Kong, China
{wkwong,wcwoo}@hko.gov.hk



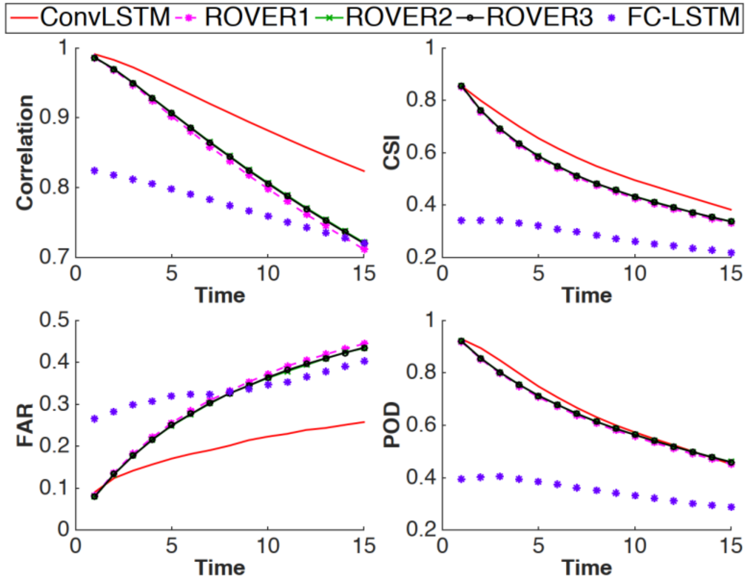
LSTM

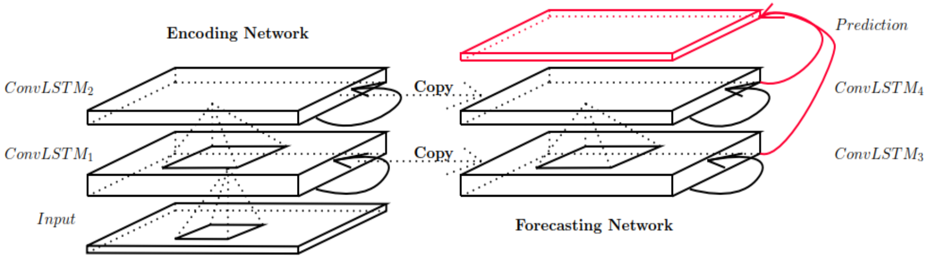


source: wikipedia

Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting

- 8148 training sequences,
- 2037 testing sequences
- 2037 validation sequences





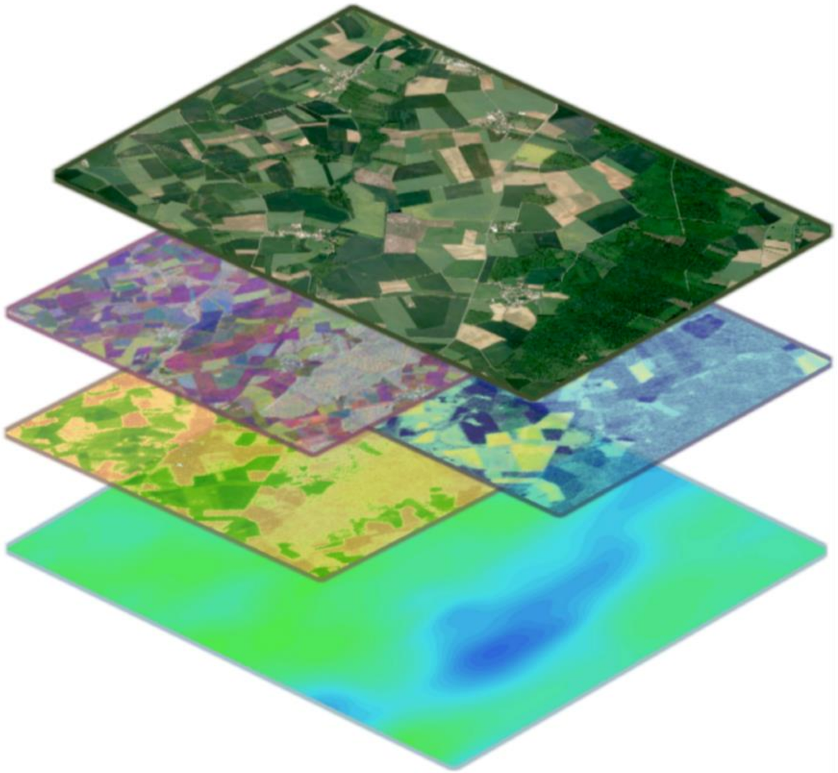
$$\mathrm{CSI} = \frac{\mathrm{hits}}{\mathrm{hits+misses+falsealarms}}$$

$$\mathrm{FAR} = \frac{\mathrm{falsealarms}}{\mathrm{hits+falsealarms}}$$

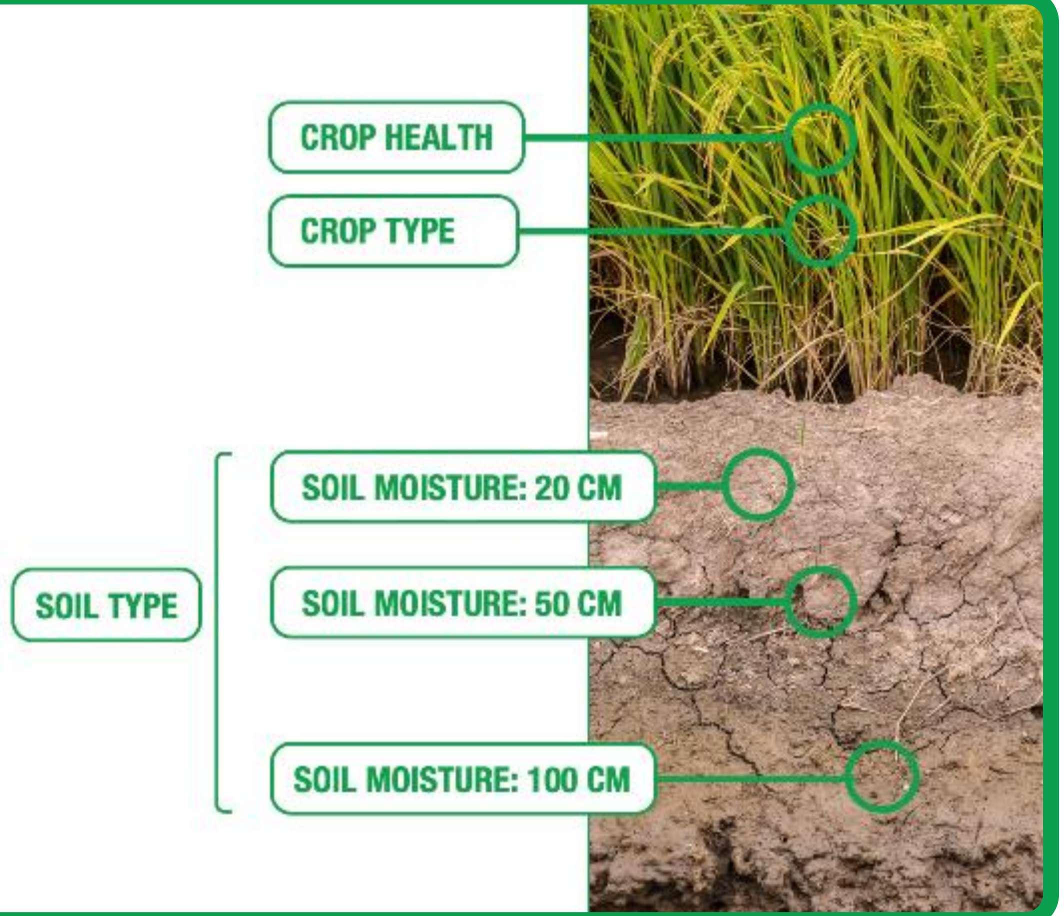
$$\mathrm{POD} = \frac{\mathrm{hits}}{\mathrm{hits+misses}}$$

Source : [Advances in Neural Information Processing Systems 28 (NIPS 2015)]

HelioPas AI



DATA FUSION & ARTIFICIAL INTELLIGENCE



HelioPas AI

With artificial intelligence we provide the needed improvement on data quality and resolution to enable a lot of use cases

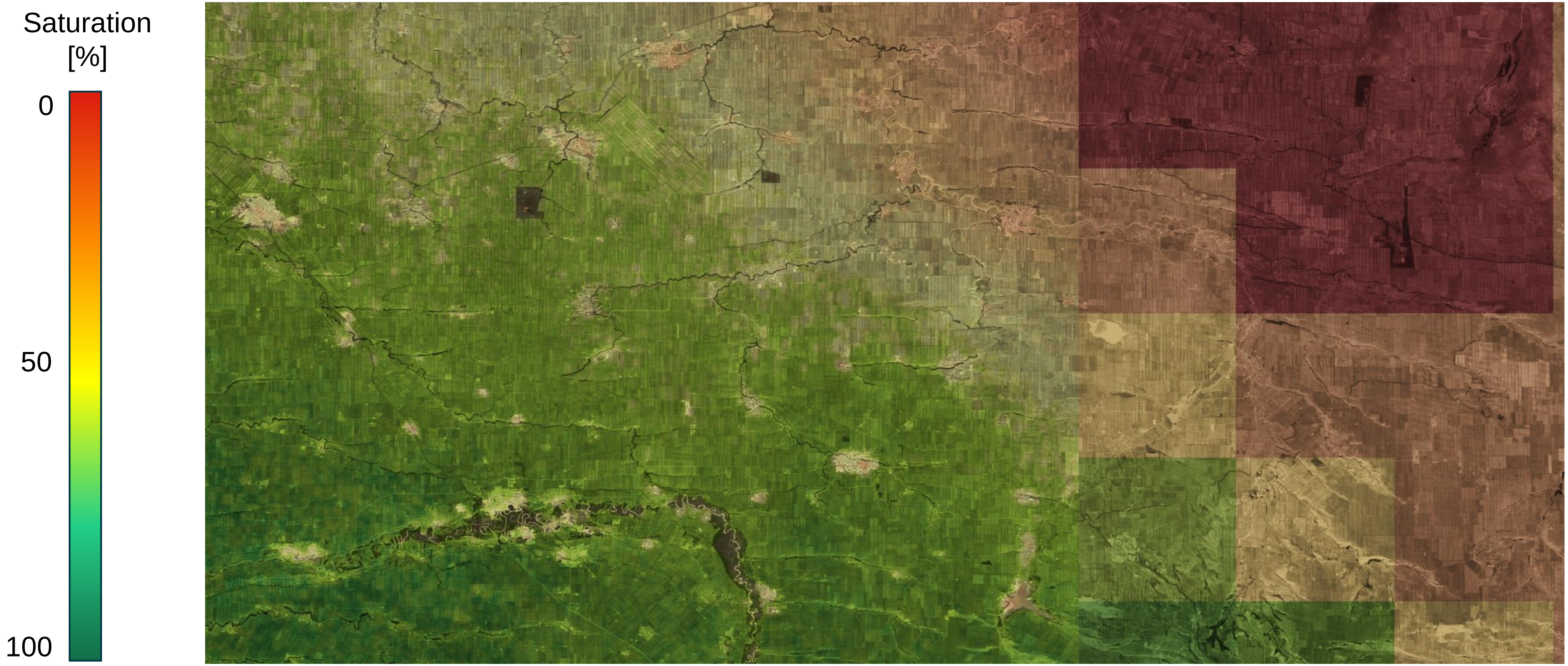
HelioPas AI

With artificial intelligence we provide the needed improvement on data quality and resolution to enable a lot of use cases

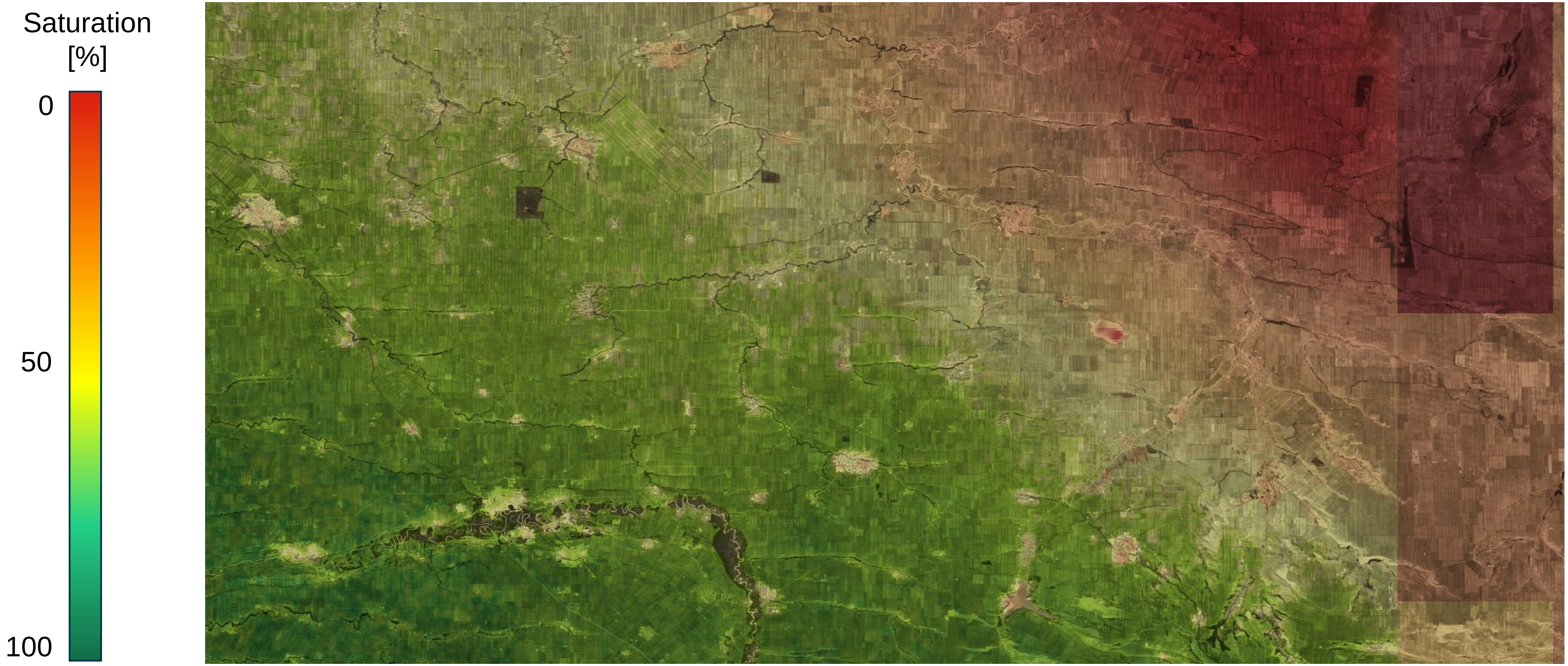
With artificial intelligence we provide the needed improvement on data quality and resolution to enable a lot of use cases

HelioPas AI

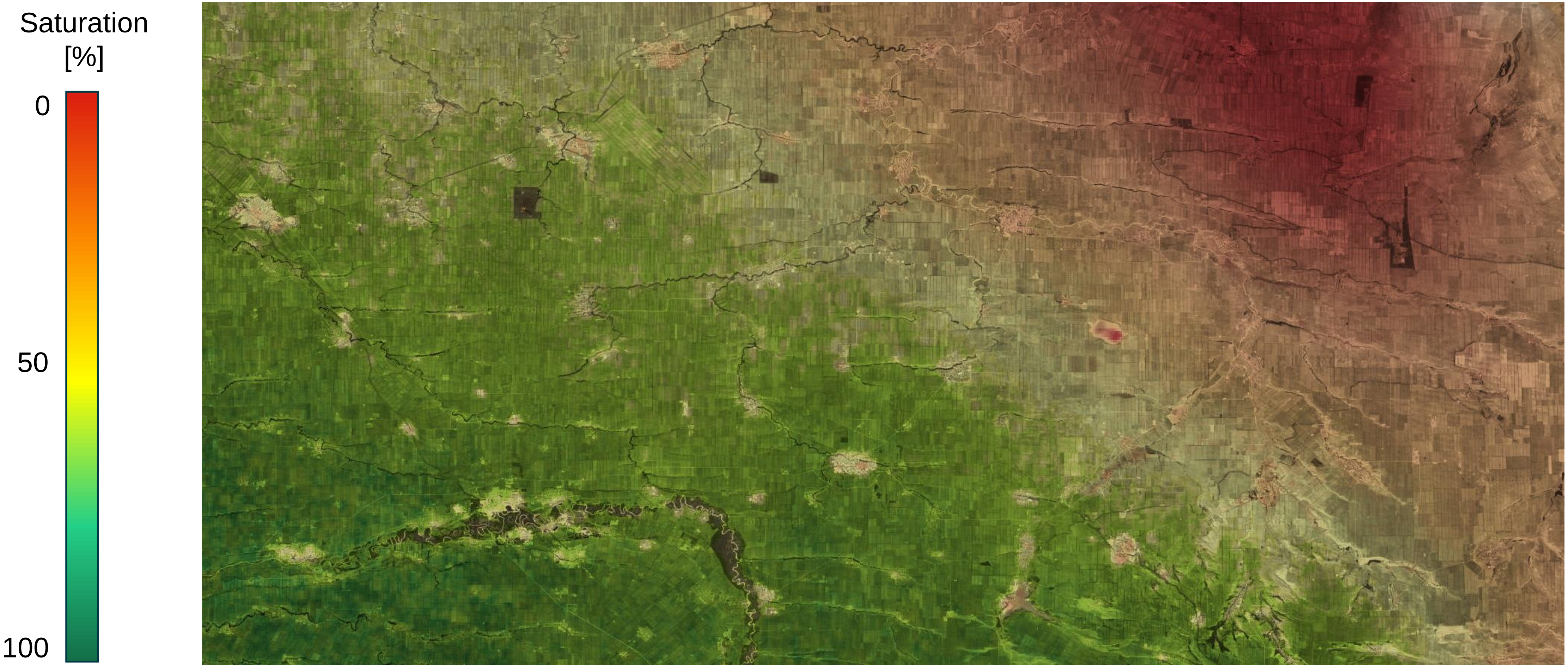
With artificial intelligence we provide the needed improvement on data quality and resolution to enable a lot of use cases

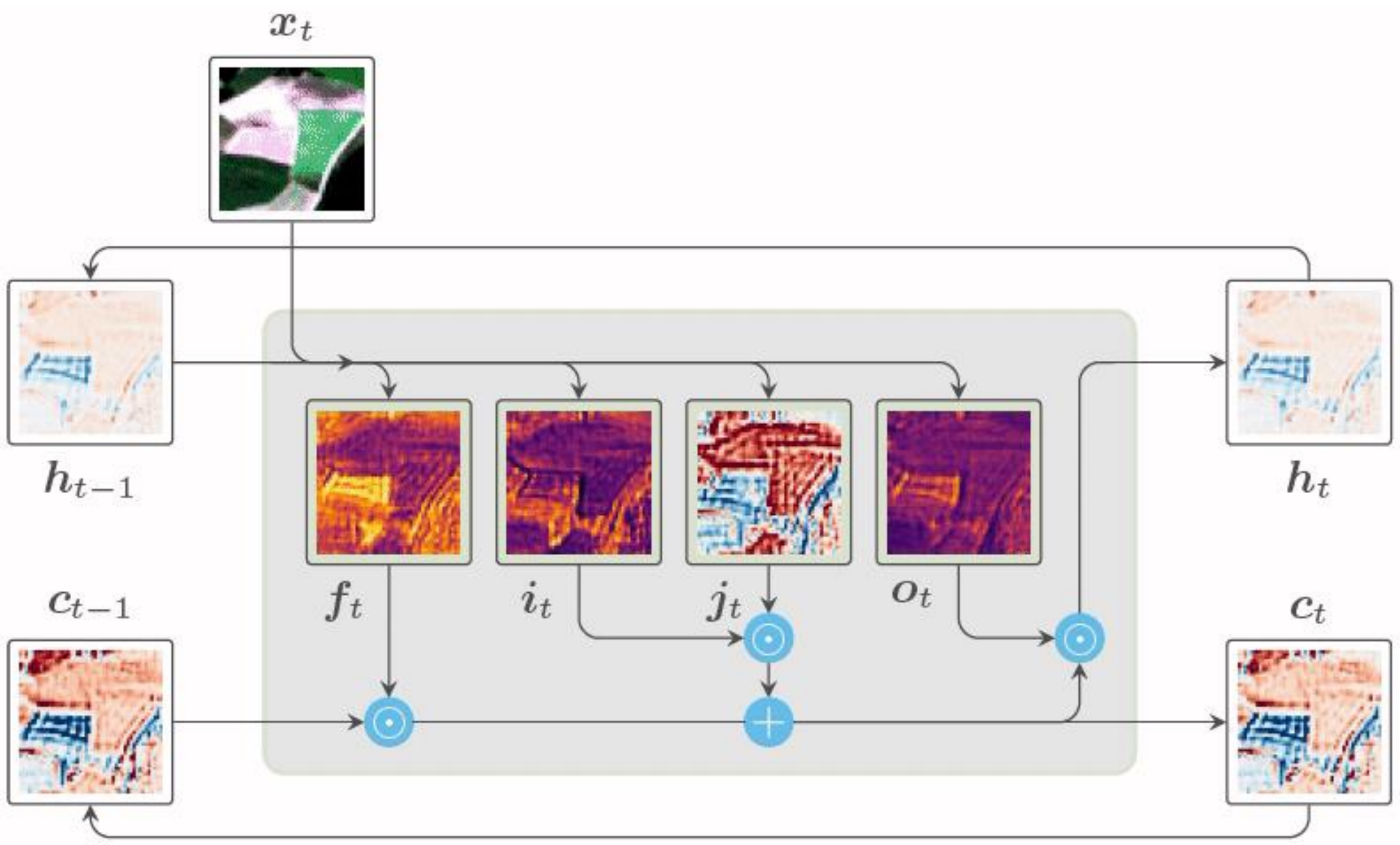


With artificial intelligence we provide the needed improvement on data quality and resolution to enable a lot of use cases



With artificial intelligence we provide the needed improvement on data quality and resolution to enable a lot of use cases




x_t
h_{t-1}
h_t
c_{t-1}
c_t
f_t
i_t
j_t
o_t

SmartAQnet

Karlsruher Institut für Technologie

- funded by the Federal Ministry of Transport and Digital Infrastructure, 04/17-03/20

Interdisciplinary and multi-method!

  - Epidemiology
  - Chemistry
  - Metereology
  - Geography
  - Computer Science

- Based on a pragmatic**, data driven approach**
- Combination of existing data sets with a **networked mobile measurement strategy**

Gefördert durch:

Bundesministerium für Verkehr und digitale Infrastruktur

mFUND Das Startkapital für die Mobilität 4.0

HelmholtzZentrum münchen
Deutsches Forschungszentrum für Gesundheit und Umwelt

Karlsruher Institut für Technologie

Universität Augsburg University

AEROSOL AKADEMIE

GRIMM DURAG GROUP

Technical goals in SmartAQnet

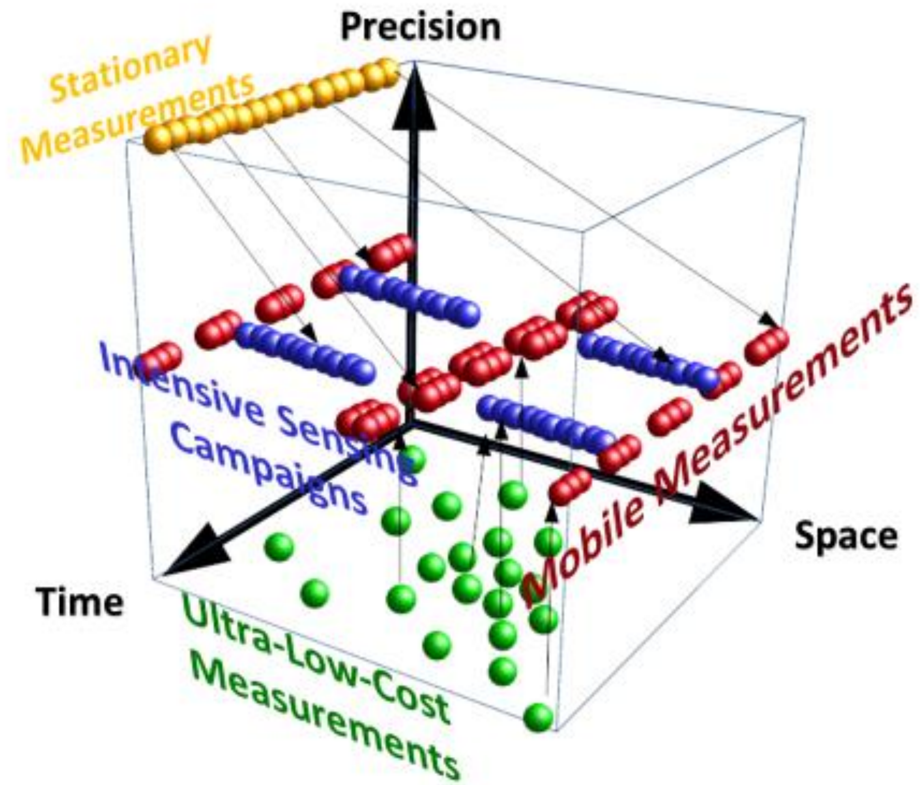
- Development of an open, participatory system for recording, visualization and prediction of spatial distribution of **air pollutants in urban atmospheres**

- Implementation of an intelligent, reproducible, finely-tuned (spatial, temporal), yet **cost-effective air quality measuring network**

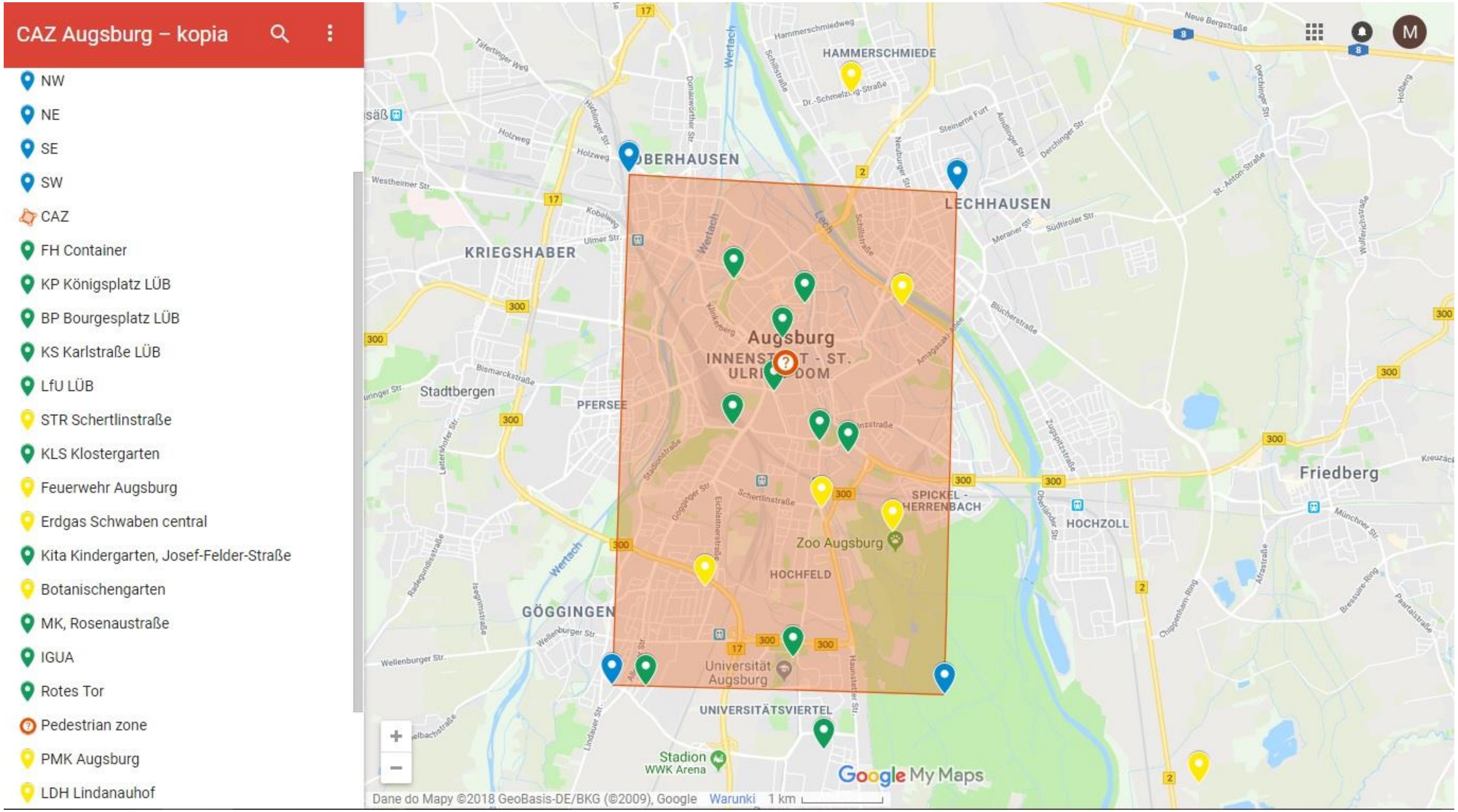
- Implementation of **small-scale numerical simulation** by GRAMM/GRAL, PALM 4U (Project Urban Climate under Change) and AUTH model chain (Moussiopoulos et al.) for determination of air pollution exposure with corresponding emission inventory (e.g. from traffic counting)

- Real-time and historical data products and applications for **science, public authorities and citizens** alike compatible for scientific purposes and user-oriented service development

Cross-calibration & Cross-validation



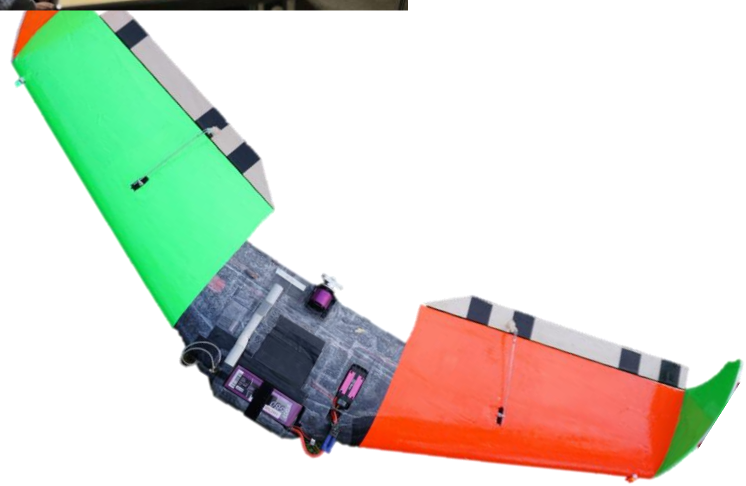
Today(!) Intensive Operation Period in Augsburg



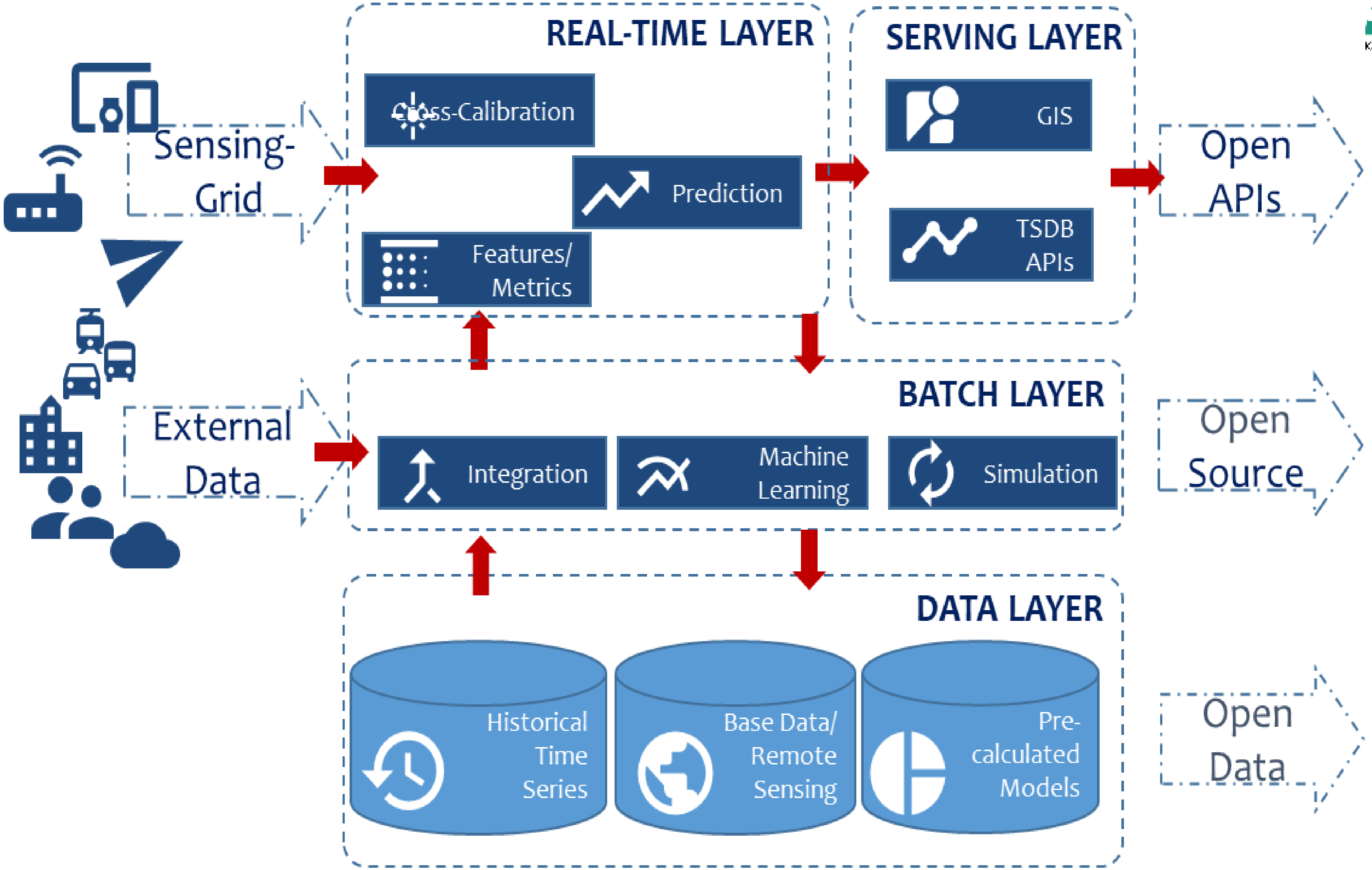
Hermann-Billing-

- Citizen Science / Participation
  - Get high resolution measurements
  - Build trust through participation
  - People do it anyways!

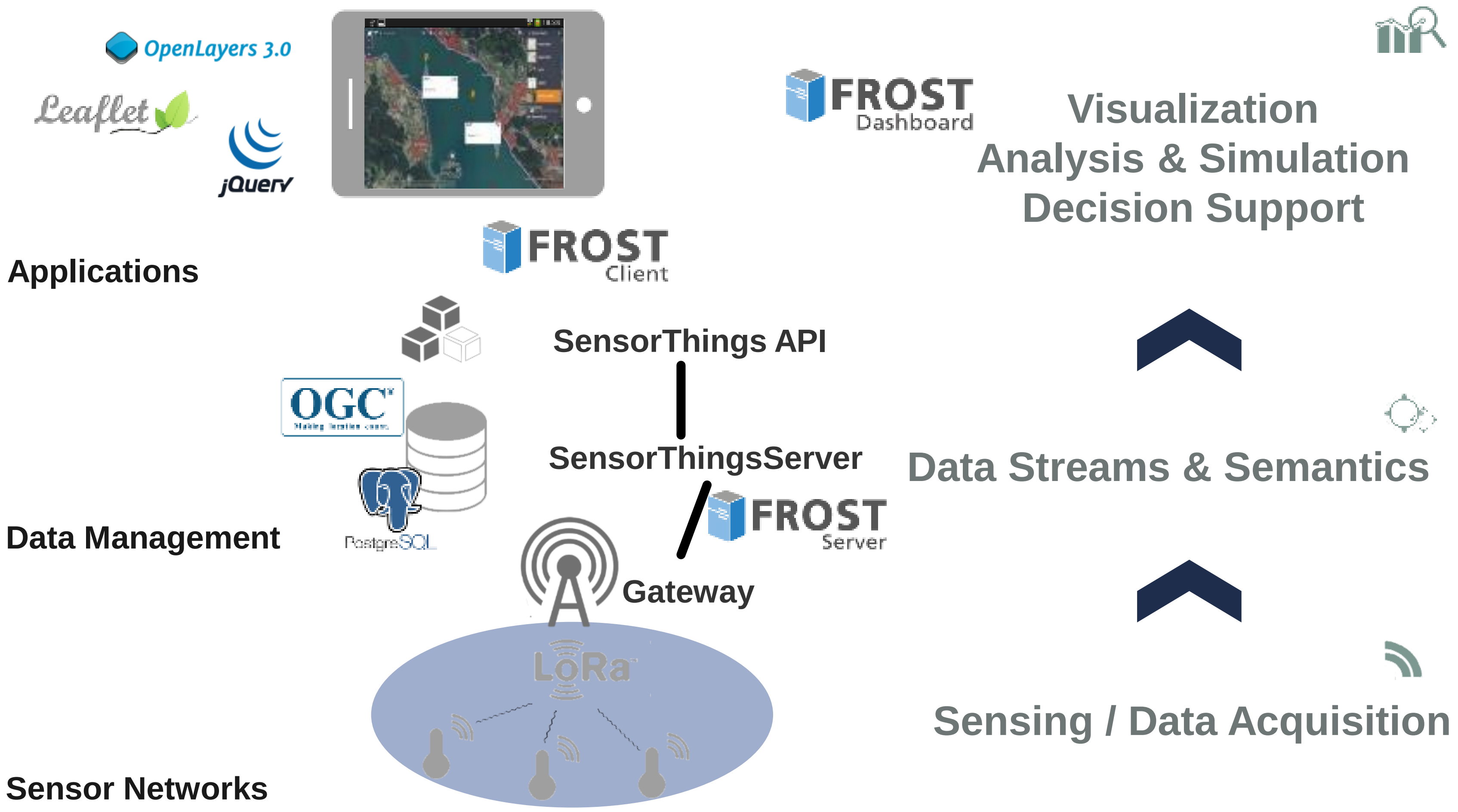
- UAV based 3D-Measurements
  - Only way to reliably detect boundary layers (cause of most extreme situations)
  - Mechanical and chemical effect insufficiently studied in 3D above cities…



- Mobile Ground measurement using e.g. bicycles but also with passive collectors
  - Mobile measurements are a key to good coverage of foreground emissions
  - Direct measurement of exposure


REAL-TIME LAYER
Cross-Calibration
Prediction
Features/ Metrics
SERVING LAYER
GIS
TSDB APIs
Sensing-Grid
External Data
Open APIs
BATCH LAYER
Integration
Machine Learning
Simulation
Open Source
DATA LAYER
Historical Time Series
Base Data/ Remote Sensing
Pre-calculated Models
Open Data

FRAUNHOFER-IOSB: STANDARDS-BASED IOT SW-ARCHITECTURE



fernando.chaves-salamanca@iosb.fraunhofer.de

The big questions towards a better picture

- **Simulation vs. Measurements**
  - Can we define apriori how much we need of both?
  - Often cross-„calibrated“ or too expensive: how can we validate both
- **Prove validity and create trust**
  - People do not trust the positioning of existing stations
  - Also the most expensive measurement station can be easily tampered
- **Scalable measurement and reuse of data**
  - Measurements implies a dedicated measurement goal
  - Cross domain use of data is not forseen in our measurement theory

Measurement vs. Simulation

- direct
- Only measurements detect systematic problems!
- Calibration is a major issue particularly for any optical system (we currently consider mass concentrations relevant)
- Spatial resolution needs more sensors: No easy way to interpolate measurements yet (this would also solve calibration)
- Relative measurements with high time resolution seems feasible!

- indirect
- We model only what we already know!
- We need fine granular input data (even a large truck can change a wind field)
- We can do scenarios!
- Models grow increasingly more complex (and thus error-prone?)
- Only computing power limits time and space resolution

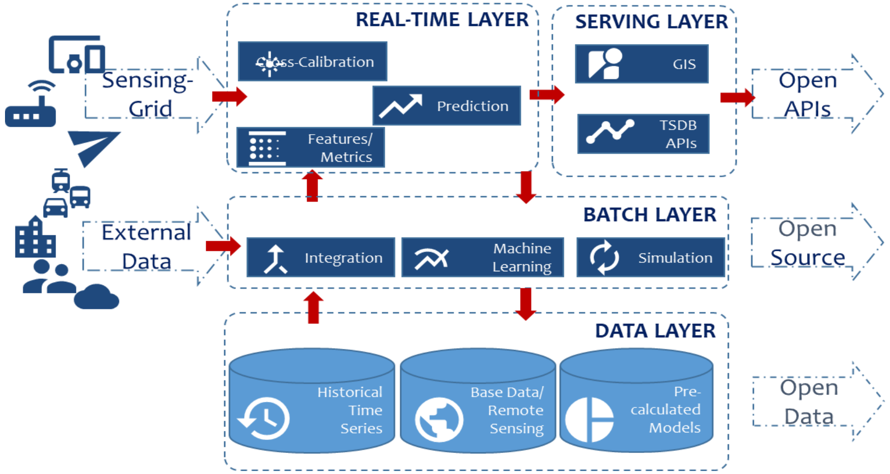
➔ Both are most of the time „not even wrong“
➔ We need more validation!!
➔ Probably we need integrated models

SmartCities: IoT, Big Data & AI

Measurement in the 21st century:
Any observation that reduces uncertainty

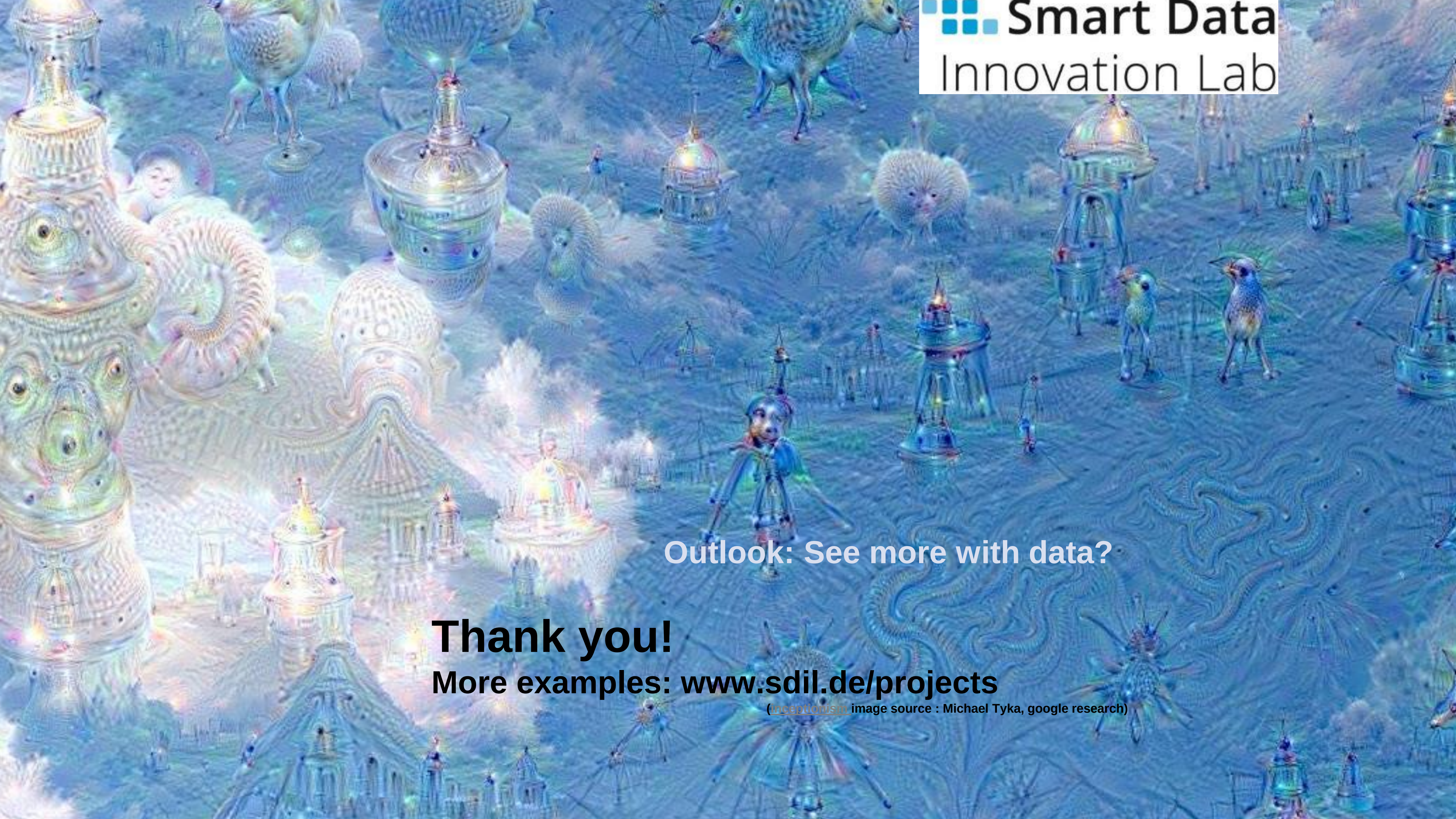
- heterogeneous data sources (variety)
- scalable data analytics (volume)
- realtime data processing (velocity)

→ For complex Smart Cities we need more **data-driven and participatory** science, journalism, decision making,…



Challenges in understandability: blackbox (**analytics**) vs. whitebox (**simulation**), „real“ **measurement** vs. simulation/prediction to understand complex systems

Exchange on **research, innovation and education** needed



Smart Data Innovation Lab

Outlook: See more with data?

Thank you!

More examples: www.sdil.de/projects

(inceptionism image source : Michael Tyka, google research)