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

COST Action TD1105

Final Meeting at PRAGUE (CZ), 5-7 October 2016

New Sensing Technologies for Air Quality Monitoring

Towards Intelligent Air Quality Monitoring Networks:

How *Machine Learning* Improve the Accuracy of Air Quality Multisensors Systems

ENEA
Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile

Saverio De Vito

WG Member & MC Substitute

ENEA / IT

saverio.devito@enea.it

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

Scientific context and objectives in the Action

Background / Problem statement:

Along this years, our participation in EuNetAir was basically motivated by:

- Explore the possible impact of *machine learning* onto developing precise, stable and accurate microsensors based AQ systems
- Explore the *feasibility of pervasive, cooperative and mobile* microsensors based air quality monitoring network

We found an <u>outstanding</u> scientific level, a

challenging, *cooperative* and *friendly*

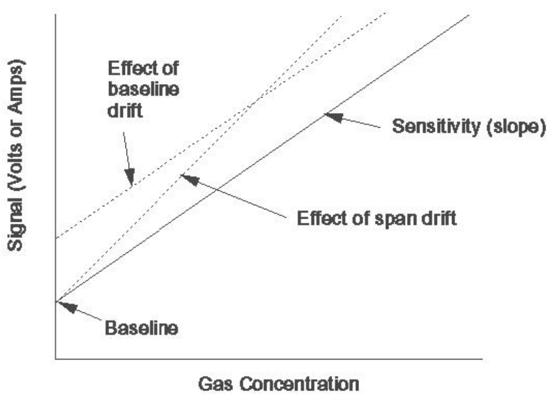
environment that supported us and gave a context to our research!

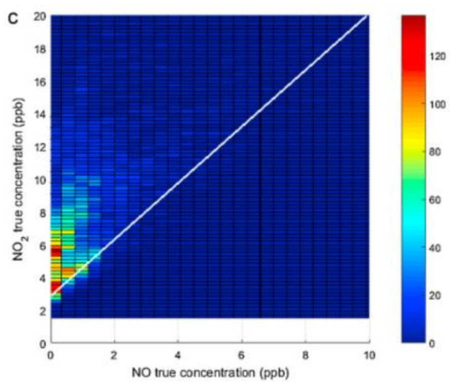
Research activities of the Partner

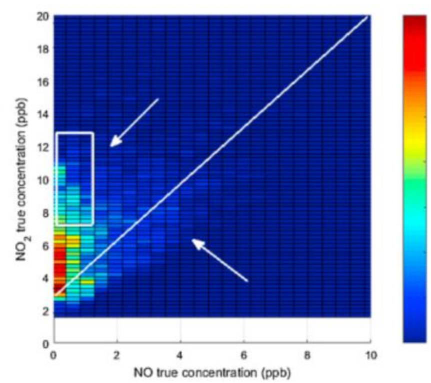
Main challenges:

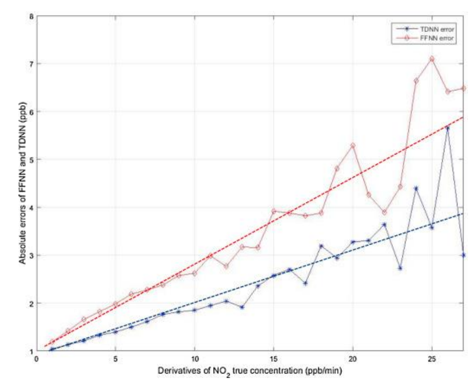
- **Sensor issues:** Cross Sensitivities, Instabilities/Drift, Slow Dynamics, Fabrication variability

- **Integration/Cost Effectiveness** of a pervasive and heterogeneous architecture





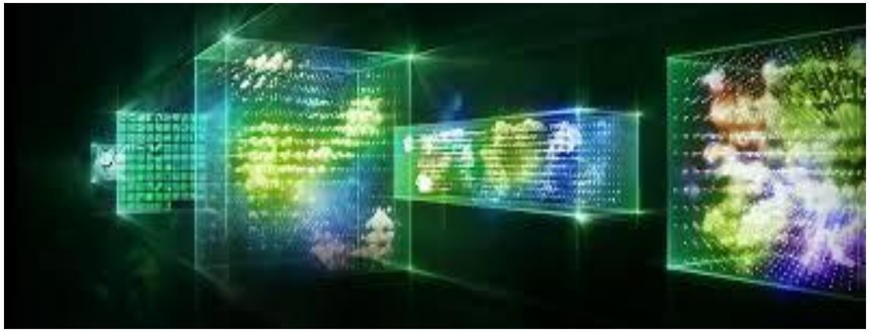


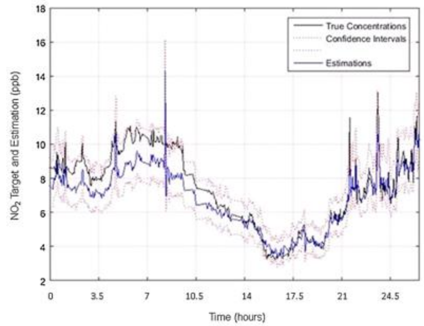


Research Activities

S/T Advancements: In these years We have gone…

- …From basic NN based solutions to:
    - *On board calibration sw implementation*
    - *Dynamic machine learning solutions*
    - *Semisupervised learning for drift counteraction*
    - *Cross calibration for drift counteraction*
    - … and now moving into the deep learning realm
- …. From lab based single networked e-nose to:
    - *Custom smartphone centered ecosystem*
    - *Cooperative air quality monitoring solution*





Machine Learning

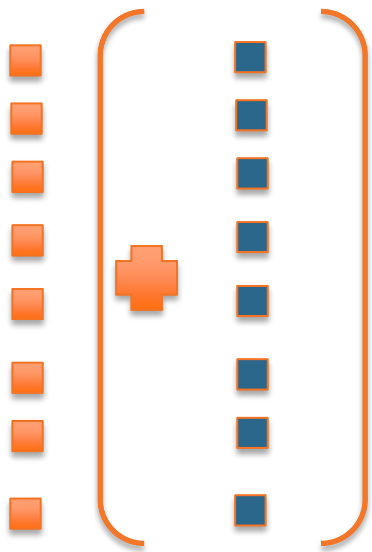
Machine learning **enables computational machines to accomplish a task (almost) without being explicitly programmed to.**

Exploits:

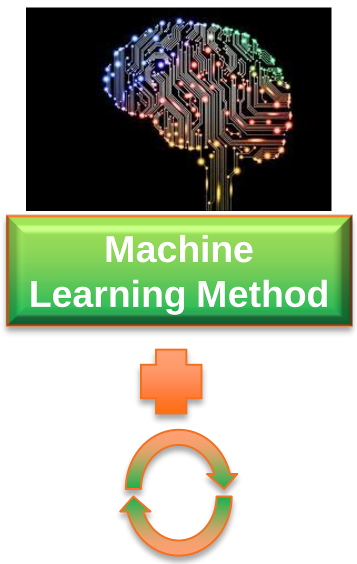
Unlabeled Data
Unsupervised Learning

Labeled Data
Supervised Learning

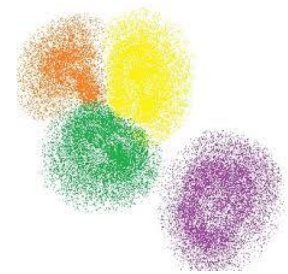
In passive or active, static or adaptive ways.



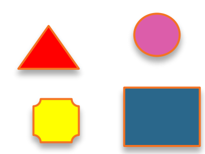
Data **Labeled Data**



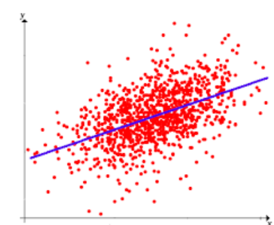
Learning Technique



Clustering Tasks



Classification Tasks



Regression Tasks

Lot of efforts expressed during these years….

Intercomparison of air quality data using principal component analysis, and forecasting of PM_{10} and $PM_{2.5}$ concentrations using artificial neural networks, in Thessaloniki and Helsinki

Dimitris Vo
Ari Karppin

[a] Department of Me
[b] Finnish Meteorolog
[c] Department of Env

Calibration of a cluster of low-cost sensors for the measurement of air pollution in ambient air

L. Spinelle[1], M. Gerboles[1], M.G. Villani[2],

Dynamic neural network architectures for on field stochastic calibration of indicativ

E. Esposito[a,*], S. De Vito[a], M.

[a] UTTP-MDB, ENEA, P.le E. Fermi, 1, 80055 Portici (
[b] Dept. of Chemistry, University of Cambridge, Lens

Field calibra
air quality m

Laurent Spinelle[a,]
Fausto Bonavitacol
Show more

What have We learned?

On-the-fly Calibration of Low-cost Gas Sensors

David
Computer Engineerin
{hasenf

A Computational Intelligence Approach to Diabetes Mellitus and Air Quality Levels in Thessaloniki, Greece

Kostas Karatzas, Vassili
Christos Savopoulos, Apostc

Contents lists available at ScienceDirect
Thin Solid Films
journal homepage: www.elsevier.com/locate/tsf
ELSEVIER

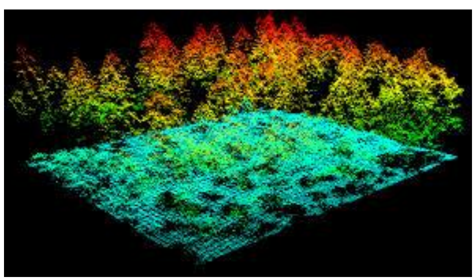
Exploring the selectivity of
naphthalene mixture using

M. Bastuck[a,b,*], D. Puglisi[b], J. Huot
M. Andersson[b,c], A. Schütze[a]

IEEE SENSORS JOURNAL, VOL. 12, NO. 11, NOVEMBER 2012 3215

Semi-Supervised Learning Techniques in Artificial Olfaction: A Novel Approach to Classification Problems and Drift Counteraction

Saverio De Vito, *Member, IEEE*, Grazia Fattoruso, Matteo Pardo, Francesco Tortorella, *Senior Member, IEEE*, and Girolamo Di Francia



Research Activities: Lesson learned

1. On Field Calibration with non linear multivariate algorithms outperform linear lab based calibration

ELSEVIER

Available online at www.sciencedirect.com

ScienceDirect

Sensors and Actuators B xxx (2007) xxx–xxx

SENSORS AND ACTUATORS B CHEMICAL

www.elsevier.com/locate/snb

On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario

S. De Vito[a,*], E. Massera[a], M. Piga[b], L. Martinotto[b], G. Di Francia[a]

[a] ENEA, Centro Ricerche Portici, 80055 Portici (NA), Italy

[b] Pirelli Labs, Viale Sarca 222, 20126 Milano, Italy

Received 8 June 2007; received in revised form 13 September 2007; accepted 14 September 2007

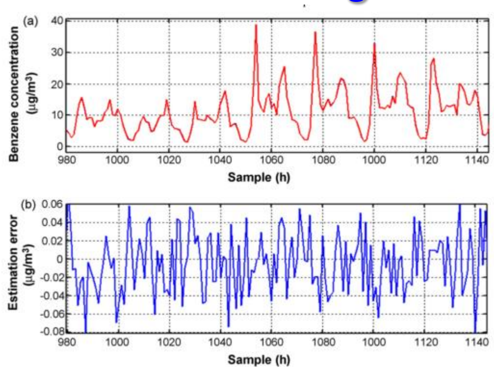
$$C_j = \Psi(\mathrm{RSens}_i)$$

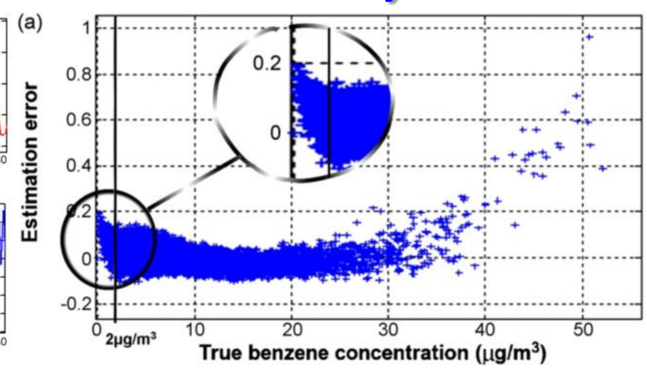
- Long term dataset (available on UCI)
- ANN regression, optimize # of calibration samples
- Warn against calibration «locality»

Table 4
Relationship between training set length and performance indexes for benzene concentration prediction

Training length	MRE	STD_RE	MAE ($\mu g/m^3$)	STD_AE ($\mu g/m^3$)	SCC
24 h	0.50	1.07	5.09	5.63	0.4383
96 h	0.16	0.24	1.31	1.65	0.8849
10 days	0.020	0.11	0.13	0.25	0.9938
25 days	0.017	0.071	0.081	0.19	0.9986
50 days	0.012	0.057	0.050	0.19	0.9984
100 days	0.009	0.076	0.050	0.16	0.9995

Results show, as expected, a definitely positive trend extending the duration of training set. Surprisingly, a 10-day training length is sufficient to achieve a 2% relative error; still, error standard deviation remains significant.





Research Activities: Lesson learned

1. On Field Calibration with non linear multivariate algorithms outperform linear lab based calibration


Sensors and Actuators B: Chemical

Volume 215, August 2015, Pages 249–257

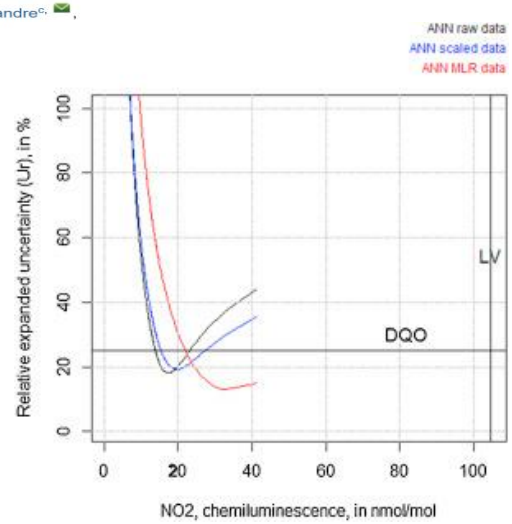
Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide

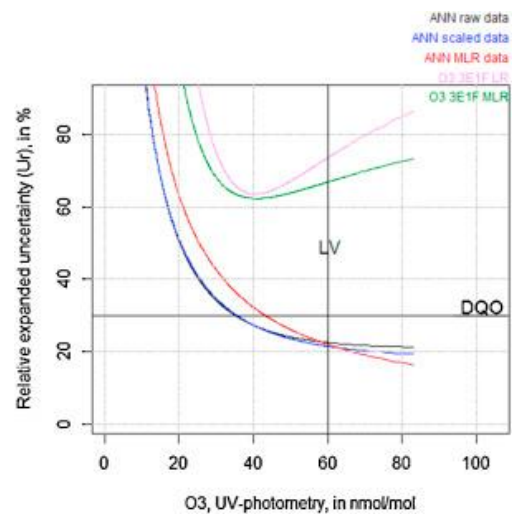
Laurent Spinelle[a], Michel Gerboles[a], Maria Gabriella Villani[b], Manuel Aleixandre[c], Fausto Bonavitacola[d]

Show more

$$C_j = \Psi(\mathrm{RSens}_i)$$

- Significant # of sensors
- Different regression methodologies





- ANN outperform LR & MLR
- Reach EC/2008 Dir DQO Level
- Solve cross sensitivity
- «Solve» *locality* by discarding T,RH information

Research Activities: Lesson learned

2. Machine learning algorithms are both efficent (frugal data needs) and effective

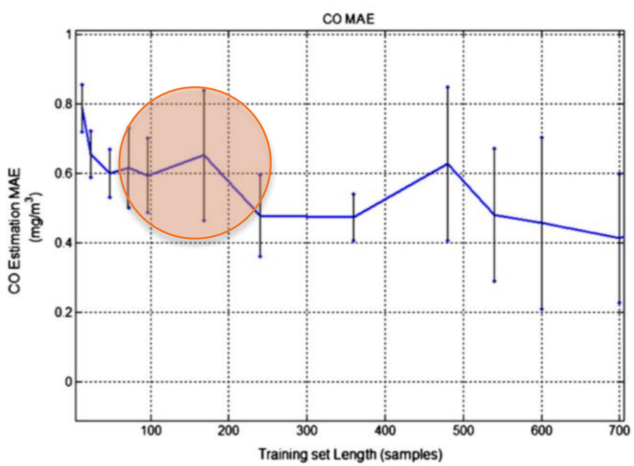


Fig. 1. CO concentration estimation MAE, expressed in mg/m^3, versus training set length measured in samples (hours) with related confidence intervals. All sensor responses have been used as feature vector (crossvalidation setting, see Table 4 for details).

Sensors and Actuators B 143 (2009) 182–191

Contents lists available at ScienceDirect

Sensors and Actuators B: Chemical

journal homepage: www.elsevier.com/locate/snb

CO, NO_2 and NO_x urban pollution monitoring with on-field calibrated electronic nose by automatic bayesian regularization

Saverio De Vito[a,*], Marco Piga[b], Luca Martinotto[b], Girolamo Di Francia[a]

[a] ENEA, Centro Ricerche Portici, 80055 Portici (NA), Italy
[b] Pirelli Labs, Viale Sarca 222, 20126 Milano, Italy

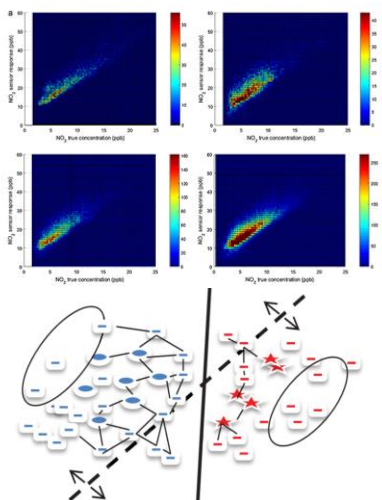
Table 6
Performance evaluation for different feature vector compositions in the NO_2 concentration estimation problem. Ten days long training set has been used. Confidence intervals are reported for MRE and MAE.

Feature set							MRE	MAE (µg/m^3)
CO	NMHC	NOx	NO_2	O_3	T	RH		
X							0.31 ± 0.07	36 ± 13.0
	X						0.29 ± 0.09	29 ± 9.8
		X					0.43 ± 0.22	33 ± 6.7
			X				0.47 ± 0.13	50 ± 24.7
				X			0.29 ± 0.07	29 ± 8.25
		X		X			0.29 ± 0.11	26 ± 6.41
			X	X			0.36 ± 0.10	38 ± 18.3
		X	X				0.37 ± 0.11	34 ± 19.1
X	X						0.29 ± 0.09	29 ± 10.05
X		X	X				0.30 ± 0.08	32 ± 14.0
		X	X	X			0.35 ± 0.09	38 ± 16.9
			X		X	X	0.39 ± 0.14	45 ± 23.2
		X			X	X	0.28 ± 0.05	25 ± 4.53
		X	X	X	X	X	0.26 ± 0.06	28 ± 6.27
X	X			X			0.28 ± 0.08	27 ± 9.95
X	X		X				0.23 ± 0.10	19 ± 4.01
X	X	X	X	X			0.22 ± 0.07	19 ± 7.97
X	X	X	X	X	X	X	0.22 ± 0.08	20 ± 4.00
X	X	X		X			0.28 ± 0.08	29 ± 14.62
X	X		X	X			0.22 ± 0.09	20 ± 5.01
X	X		X	X	X	X	0.22 ± 0.09	20 ± 4.35

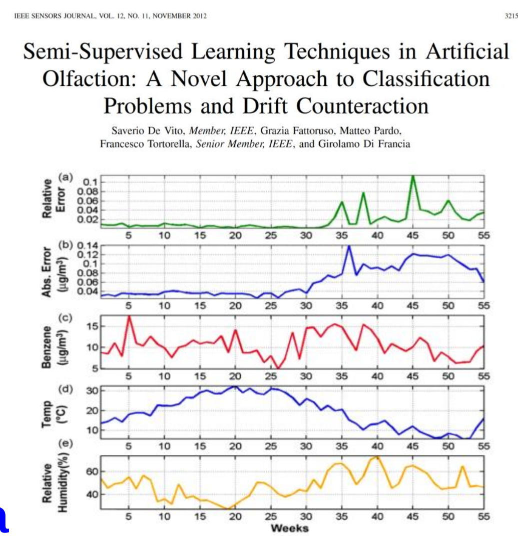
- You can achieve good performance with 1 week training set (24 h with SSL!)
- You may not need many sensors to achieve good cross sensitivities coverage
- Surprisingly You may not need information on T, RH

Research Activities: Lesson learned

3. Machine learning algorithms can be effective to counter drift



Learning from unlabeled data



IEEE SENSORS JOURNAL, VOL. 12, NO. 11, NOVEMBER 2012 3215

Semi-Supervised Learning Techniques in Artificial Olfaction: A Novel Approach to Classification Problems and Drift Counteraction

Saverio De Vito, *Member, IEEE*, Grazia Fattoruso, Matteo Pardo, Francesco Tortorella, *Senior Member, IEEE*, and Girolamo Di Francia

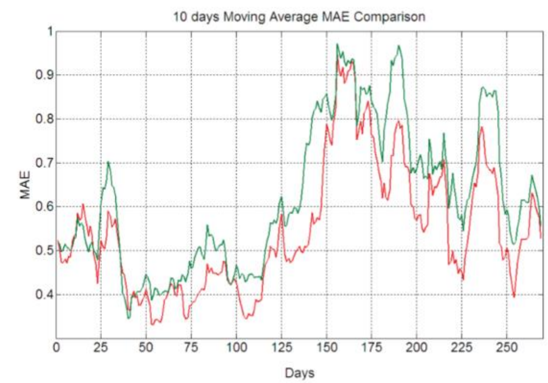


Fig. 9. Drift counter action experiment. CO Estimation comparison with integrated k-NN-BPNN SSL algorithm (red, 100 unlabeled samples reservoir) and standard NN algorithm (green) based on 24 samples. The SSL approach achieved a performance gain of 11.5% with respect to the one-year long averaged MAE score.

You can partially adapt your learnt knowledge without actually resorting to labeled data (after an initial supervised training)

Research Activities: Lesson learned

4. Dynamic Machine learning can significantly improve the responsiveness

Dynamic neural network architectures for on field stochastic calibration of indicative low cost air quality sensing systems

E. Esposito[a,*], S. De Vito[a], M. Salvato[a], V. Bright[b], R.L. Jones[b], O. Popoola[b]

[a] UTTP-MDB, ENEA, P.le E. Fermi, 1, 80055 Portici (NA), Italy
[b] Dept. of Chemistry, University of Cambridge, Lensfield Rd., Cambridge, UK

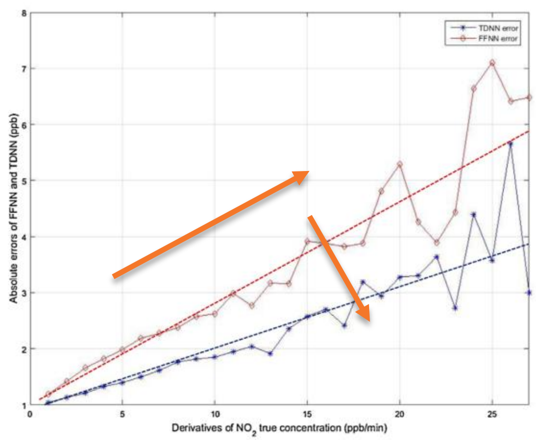
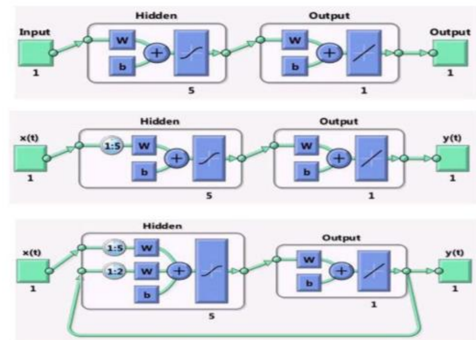
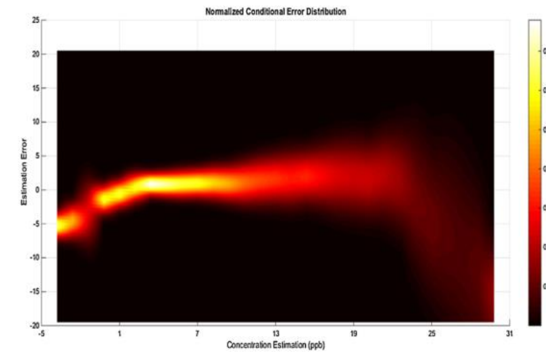
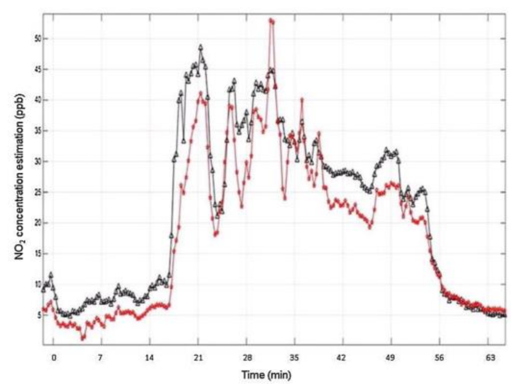
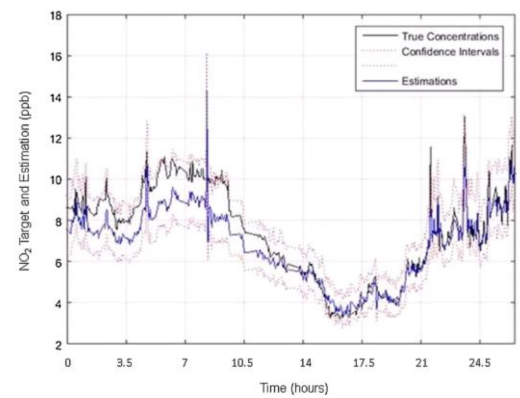


Fig. 10. Comparison of average absolute error distribution trends given the absolute derivative of the NO_2 target gas concentration (ppb/min) for dynamic (below) and static network. The TDNN network is capable of achieving a significant reduction of the estimation error wrt the static FFNN network. This behavior is enhanced when the target gas concentration changes quickly and high values of absolute derivative can be measured.









Research Activities: Lesson learned

5. Active/continuous learning can be exploited continously upgrade the calibration possibly starting from a coarse one.

On-the-fly Calibration of Low-cost Gas Sensors

David Hasenfratz, Olga Saukh, and Lothar Thiele

Computer Engineering and Networks Laboratory, ETH Zurich, Zurich, Switzerland
{hasenfratz,saukh,thiele}@tik.ee.ethz.ch

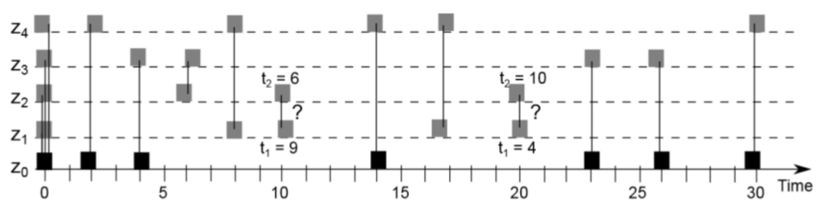


Fig. 5. Multi-hop calibration. *Connected sensors are in each others temporal and spatial vicinity.*

Data can come from (temporarily) colocated sensors
But possibly from geospatial models

Table 1. The total amount of time while the calibrated sensors measured more than 60 ppb (left) and their relative errors when compared to the 66.2 h measured by the fixed station (right).

Calibration	Initial	Forward	Backward	Instant	Initial	Forward	Backward	Instant
Sensor 1	90.5 h	54.3 h	65.8 h	64.6 h	36.7 %	18.4 %	0.6 %	2.4 %
Sensor 2	231.3 h	69.5 h	65.6 h	69.3 h	249.4 %	5.0 %	0.8 %	4.7 %

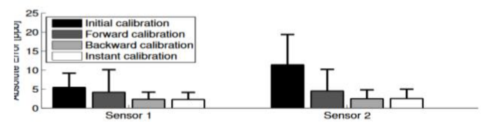


Fig. 9. Mean and standard deviation of the absolute measurement errors of the two gas sensors.

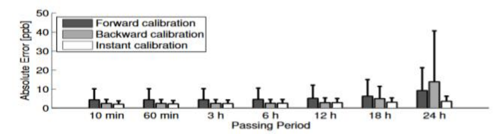


Fig. 10. Calibration accuracy for passing periods between 10 minutes and 1 day.

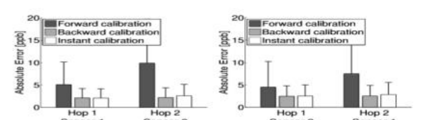


Fig. 11. Absolute measurement errors when sensors are calibrated over one and two hops.

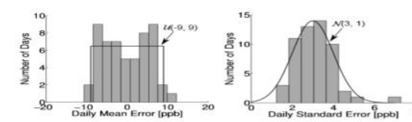


Fig. 12. Mean and standard deviation of the daily average measurement error from sensor 1.

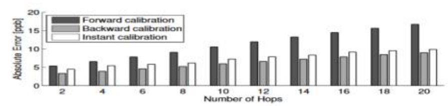


Fig. 13. Simulative analysis of the measurement errors when calibrating over multiple hops.

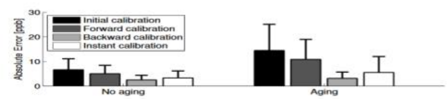


Fig. 14. Simulative analysis of the influence of sensor aging on the calibraiton accuracy.

The Vision :

Towards intelligent air quality networks

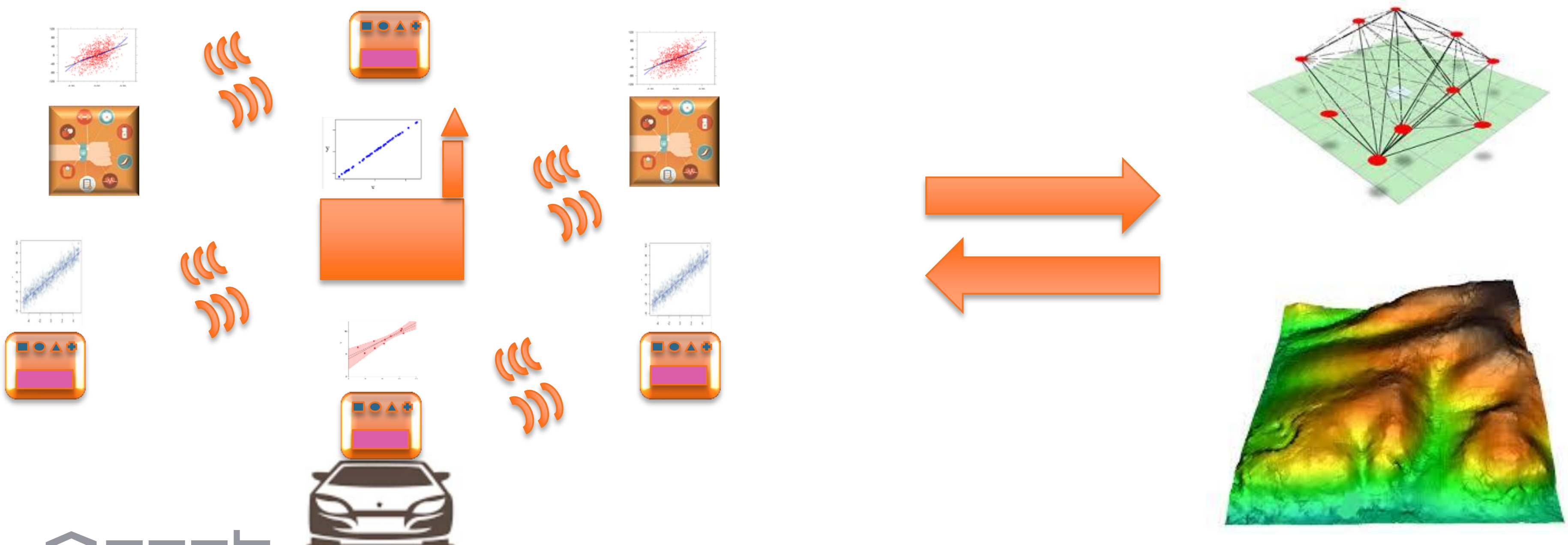
Integrating conventional and microsensors based systems.

Integrating fixed and mobile (wearable) systems…

…with different accuracy and precision…

…interacting each other to cross calibrate themselves

…and with geostatistical models for data assimilation and recalibration



Final considerations

- *Machine learning has shown as transformative and probably the best tool to analyze and make sense of chemical sensors data*

- *Its potential impacts are significantly wider that just calibration. Moreover, DBNs promises to significantly ameliorate robustness and can give us better insights on sensor data*

- *Still integration with model based approach (data assimilation) is still to be further developed*

- *<u>The amount of available data is now HUGE. A lot of open points that can be addressed now!</u>*

Thanks you for your attention…

Your friendship…

And really fruitful scientific cooperation!

Thank you Michele…

For your tremendous effort!

saverio.devito@enea.it

Special Session #4

Algorithms for distributed chemical sensors

- **Calibration Algorithms for pervasive/mobile units**
- **Cooperative sensing platforms and algorithms**
- **Data fusion algorithms from pervasive units**
- **Data assimilation approaches**

Welcome to
ISOEN 2017
Registration available soon

Important Deadlines:

Special session & tutorial proposals	November 6, 2016
Notification of special session & tutorial acceptance	December 4, 2016
Submission of regular papers	January 15, 2017
Notification of paper acceptance	March 15, 2017
Author registration	March 20, 2017
Camera ready uploads	March 28, 2017