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

COST Action TD1105

2nd Action Workshop, Brindisi, March 2014

Computational intelligence approach to pervasive chemical sensing challenges

S. De Vito, E. Massera, G. Fattoruso, G. Di Francia

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

Presenting Author:

Saverio De Vito, Ph.D.

(WG2/SIG2 Member, Sub. MC Member)

ENEA / Italy – saverio.devito@enea.it

EUROPEAN COOPERATION IN SCIENCE AND TECHNOLOGY

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

COST is supported by the EU Framework Programme

ESF provides the COST Office through a European Commission contract

Scientific context and objectives in the Action

Computational Intelligence in Modern Air Quality

- **Research Goal ->** To Address:

1. Specificity, Stability, Calibration, Energy Management, Deployment Issues with *Computational intelligence* (Statistical Regression Learning, Evolutionary Computing, Pattern recognition)
2. Develop Integrated Sensors/Model as a service architectures capable to cope with distributed social sensing needs

- **Within WG2 objectives:**

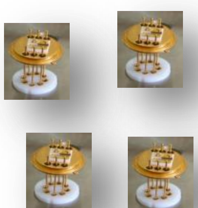
Develop integrated intelligence for networked AQC gas sensors

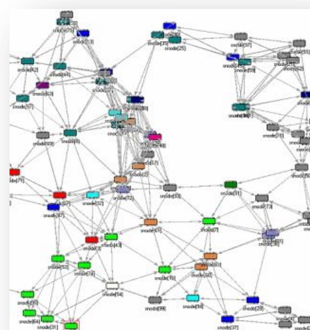
A Common framework:

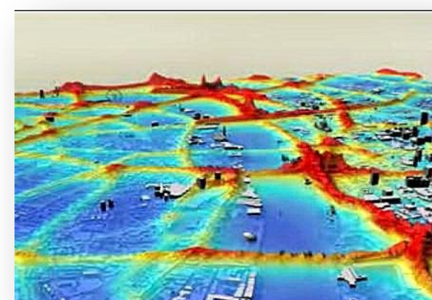
Ideally, We aim to build:

Compact-Intelligent-Cooperating-Easy-to-Deploy
Air quality chemical sensing platforms

capable to act as a network to assess air quality in complex environments







Sense… Calibrate… Cooperate… Semantic Value

A COMMON FRAMEWORK - > COMMON CHALLENGES

In this common framework, a small number of important issues seems to recurrently arise:

- We need Low Cost sensors, Low cost platforms in order to deal with pervasive deployment (stability-sensitivity-specificity trade offs)
- Effective Module Calibration (to cope with non linearity and unspecificities)
- Drift counteraction techniques (to cope with non instability)
- Calibration Transfer (to deal with sensors diversity, and numerosity)
- Energy Efficiency (Operation on Batteries)
- Sensor Fusion, Data Mining, Model integration -> (to reach high valued situational awareness)

Our Approach is to explore the possibility of *computational intelligence* techniques to reduce the impact of these issues.

Computation Intelligence

CI: *nature-inspired* computational methodologies and approaches to address complex real-world problems to which traditional approaches, i.e., first principles modeling or explicit statistical modeling, are **ineffective or infeasible.**

- Artificial Neural Networks (SVMs, FFNNs,RBFNs, etc.)

- Evolutionary computing (Genetic Algorithms, Artificial Immune Systems, Swarm,etc.)
- Fuzzy and multivalues Logics

So, What can be done for Air Quality monitoring with CI?

Computation Intelligence

It turns out that there are a number of problems with wich CI techniques can be of help:

- Multivariate non linear calibration
- Adaptive drift correction
- Optimal node localization
- Node energy efficency (tasking, censoring)
- Data mining (understanding data variance)

Computation Intelligence

It turns out that there are a number of problems with wich CI techniques can be of help:

- Multivariate non linear calibration
- Adaptive drift correction
- Optimal node localization
- Node energy efficency (tasking, censoring)
- Data mining (understanding data variance)

Multivariate non-linear calibration

Usually we know $R_{sens} = f(Conc_{TargetGas})$ + interferents list

Sometimes we could have access to:

$$R_{sens} = f(Conc_{GasA}, Conc_{GasB}, \ldots Conc_{GasN})$$

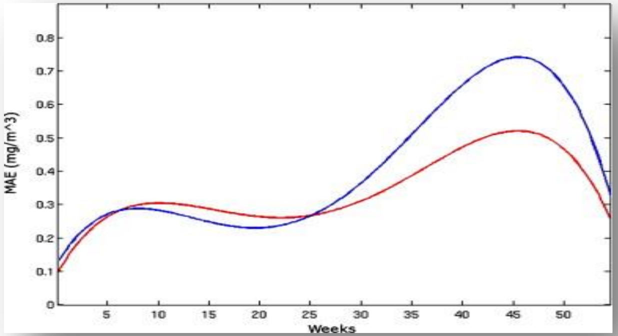
But what we really need is:

$$Conc_{GasA} = g(R_{sens1}, \ldots R_{sens2}, \ldots, R_{sensk}, T, RH)$$

Usually f and g are non linear and change over time (drift)

CI offer several solution for supervised tuning of universal non linear function approximators (e.g. ANNs or SVMs)

DRIFT Counteraction (& Training Dataset reduction)



- Sensors Drift is a well known problem for solid state based devices…
- Concept drift, often neglected, is the sensor response variation due to target variables pdf and environmental settings variation (RH, Humidity, changes in absolute and relative concentration of chemicals and their interferents, etc.)

Drift is often tackled with **recalibrations** or sensor response **correction** approaches with very interesting results.

Both these approaches require a valuable resource: **Time (=Samples)!**

- Time to calibrate the drift correction approach
- Time to recalibrate (when on field you need a GT generator!)

The Idea: Exploit **Semisupervised learning approaches** for sensors and concept drift effects reduction

Optimal node placement and tasking

In case of fixed deployments, where do I have to place my nodes?

How many nodes do i really need?

I need them to always be switched on and/or sending data?

Basically we need to find an optimal solution for a multiobjective function that should

1. *maximize our spatial accuracy*
2. *minimize the cost of the solution (deployment, mainteinance)*
3. *maximize the battery lifetime of nodes*

Evolutionary computing can offer reliable solutions with:

GAs, Swarm optimization, ANT optimization and the like.

SINGLE MODULE CALIBRATION

HOW TO TRAIN YOUR PLATFORM TO OPERATE ON FIELD

Despite cross interference?

The Goal: Calibration of a wireless (3G) 5 MOX based PV powered multisensor system for densifying air pollution monitoring network in cities.

- **Single analytes Calibration ?**

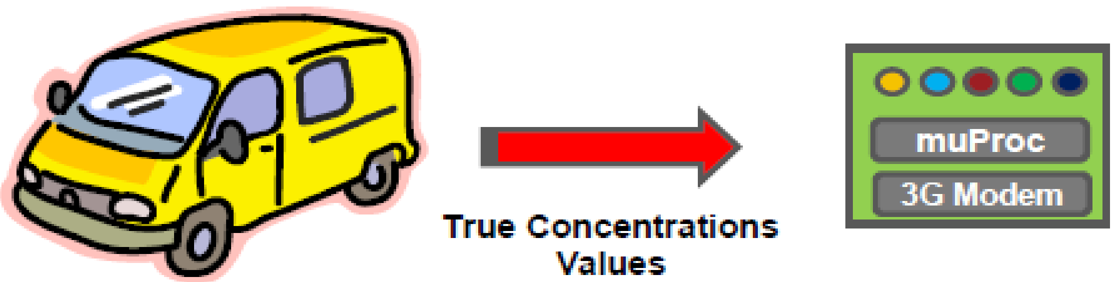
Interference set in! When mixed, gases affect the response of all sensors in your array.

- **Synthetic Mixtures ?**

You have no means to cope with the number of possible, unforeseeable interferents

The Idea: On Field Calibration

Use of a mobile spectrometers-based station to produce the GT for the statistica
Multivariate calibration (ANN, SVR) of the multisensor system responses.



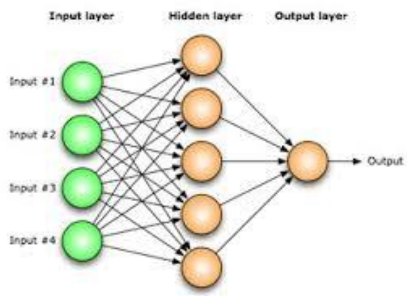
S. De Vito et al.; Sensors & Actuators,B Vol. 129, 1, Feb. 2008

SINGLE MODULE CALIBRATION:
How to train your platform to operate on Field
Building a CI model (ANN) to cope with non-specificity

We don't try to simulate what it will happen in the real world, we actually exploit what happen.

- By using gt data and sensor array response to train a batch of ANN models (one each pollutant)

Sensor Response + Rh,T



Pollutant Concentration

+ No need to synthetically generate a number of different gas mixture to obtain a valid calibration
+ In case the model can be trained also with in lab data (or can integrate the data)
+ Computationally effective

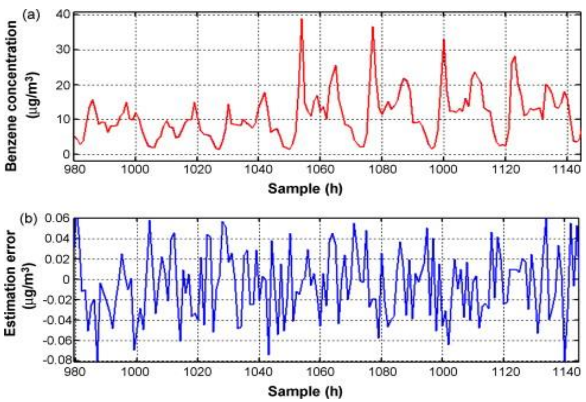
- The chosen calibration site should be representative of significant variation of pollution levels
- How long should we train? The span of time for training dataset should be both limited and representative

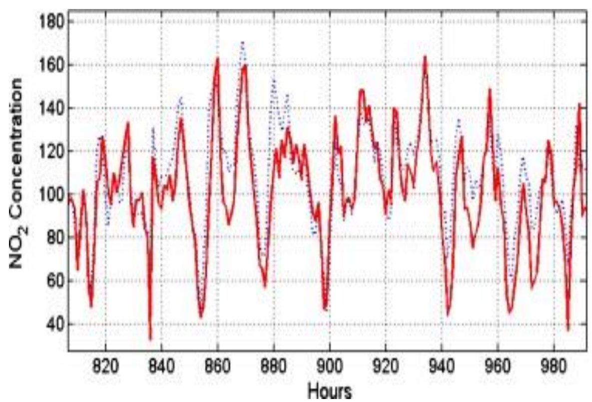
S. De Vito et al.; Sensors & Actuators,B Vol. 143, 1, Dec. 2009

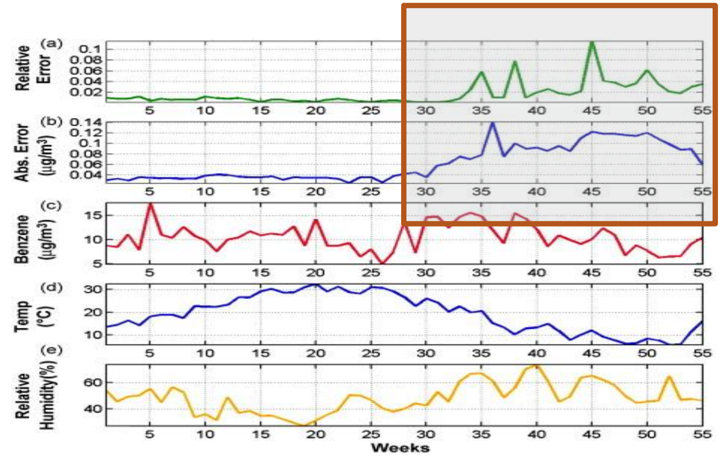
SINGLE MODULE CALIBRATION:
How to train your platform to operate on Field
Building a CI model (ANN) to cope with non-specificity

Outcome:

- ANN also learns to exploit correlations among multiple sensors [Strength but also weakness]
- Good results, **very low relative error** on the concentration estimation of Benzene and CO
- Acceptable results for the concentration estimation of NOx ,NO2 performance needs definitely to be improved







Big Issues:

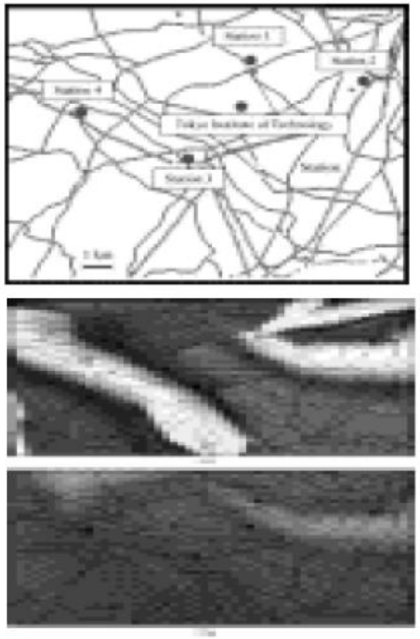
- **# of needed training samples (ten days) was too big to calibrate tenth or hundreds of multisensor devices**
- **Sensors and Concept Drift problems become significant after 4-6 Months**

The use of on-line learning (i.e. periodic recalibration) can obtain significant reduction of drift effects

S. De Vito et al.; Sensors & Actuators,B Vol. 143, 1, Dec. 2009

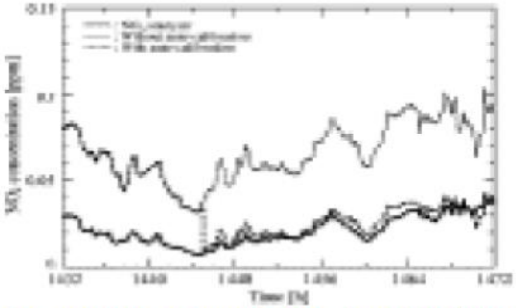
On Line Learning (Continuous periodic recalibration)

1.Based on a model:



When very low pollution levels are detected together with some favourable meteo conditions (T,RH,Wind speed) than baseline response is re-calibrated.

The procedure helped to reduce sensor drift effects.



Tsujita et al. Gas sensor network for air-pollution monitorin

Sens & Act. B, 110, 2, 200

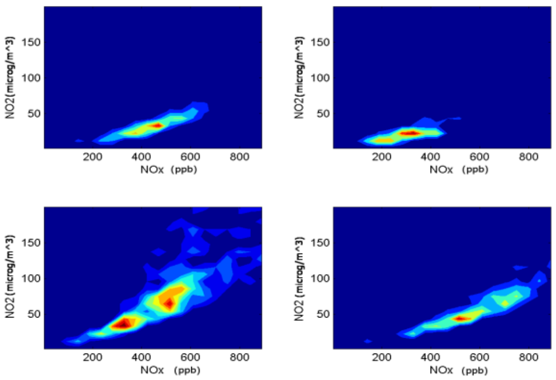
2. Based on mobile analyzers

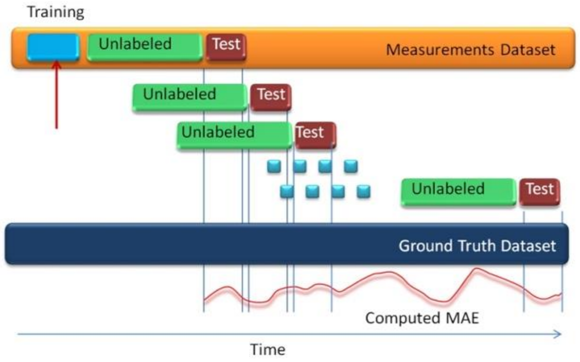
DRIFT Counteraction (& Training Dataset reduction)

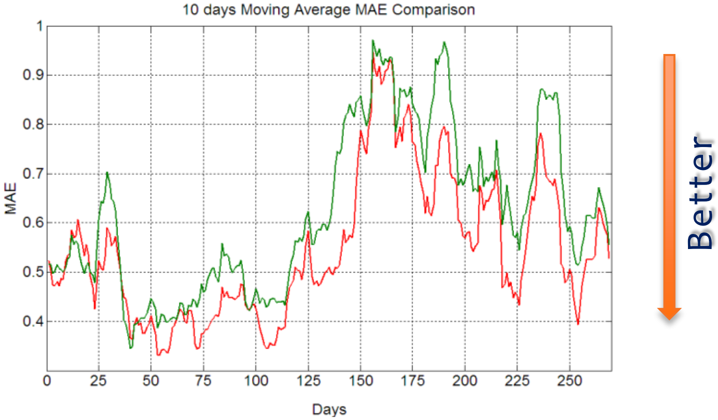
Semi supervised learning, based on manifold and cluster hypothesis, aims to exploit both

- supervised training samples (for achieving a limited but well fond knowledge of the problem)
- Unsupervised training samples to **adapt** and complete the (limited) knowledge the system has gained before

Our group applied this technique (Co-training) to the drift effect reduction in the previous setting obtaining encouraging results by using a very limited number of supervised calibration points (24Hrs).





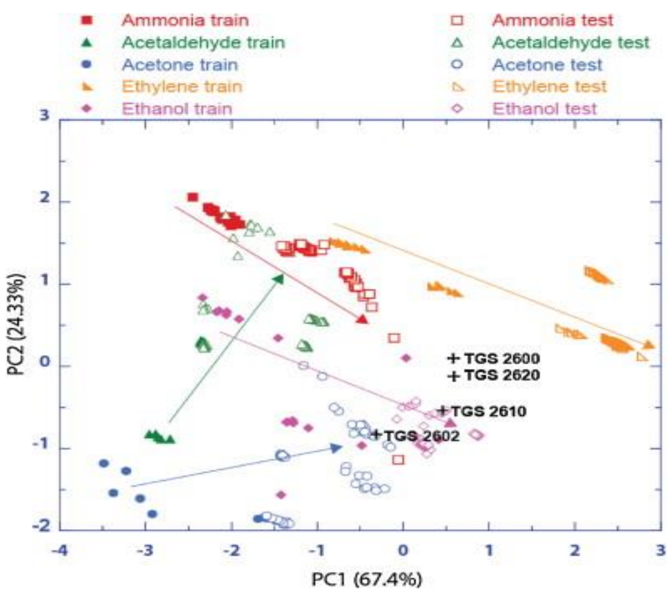


S. De Vito et al.; IEEE Sensors 2012

DRIFT Counteraction for indoor air pollutant classification

Biocircuits Institute at the University of California San Diego Dataset
-4 MOX sensors Array: Figaro Inc.: TGS2600, TGS2602, TGS2610, TGS2620
-Discrimination of ***acetaldehyde, acetone, ammonia, ethanol, and ethylene***
- 5 classes classification problem in controlled in lab conditions at constant T, RH
-509 measurements collected over a period of 18 months (HEAVY DRIFT EFFECTS)

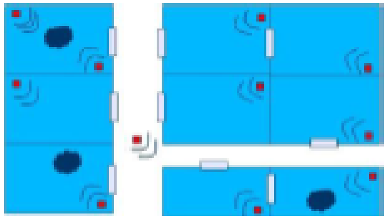
AINET^2 a new AIS (instance based classifiers) for adaptive drift correction



AINET^2 follow the variation of class centroids by semisupervised learning obtaining significant reduction of drift induced performance hits

Experimental dataset	Classification rates (%)
PLS-DA	73.73
k-NN	81.34
Supervised AINET	81.8
A^{2}INET (D = 0)	84.19 (σ = 0.92)
A^{2}INET best	95.30 (σ = 0.38) with D = 0.7

THE POWER BOTTLENECK (d-IAQ scenario)



The development of WCSN is currently hampered by technological limits on solid state sensors power management.

e.g: Most commercially MOX sensors consumes up to 400mW in their operating phase, their use is totally prevented in battery operated e-noses

Solutions:

a) Develop (RT/LT) operating sensors with good sensitivity and low LOD
b) Operate MOX with extremely low power Temp management cycles (Flammini et al,2007)

However, even when goal is reached, transmission power needs limit the operative life of continuously sampling motes (safety or security critical applications).

Sensor censoring strategies have to be developed in order to solve this issue.

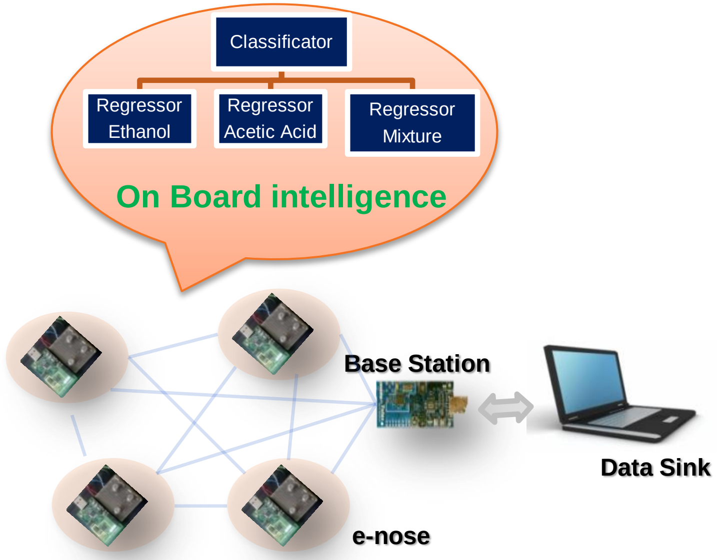
Censoring = Eliminate uninformative data transmission

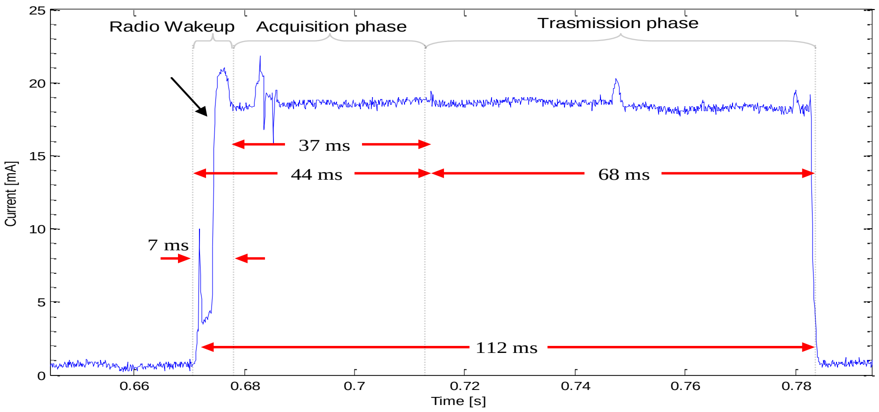
ENEA
Agenzia nazionale per le nuove tecnologie, l'energia e lo sviluppo economico sostenibile

cost
EUROPEAN COOPERATION IN SCIENCE AND TECHNOLOGY

On Board Intelligence for Sensor Censoring

The problem: Recognize uninformative data acquisitions (low concentrations of relevant pollutant or dangerous gas) in presence of interferents in a continuous monitoring scenario





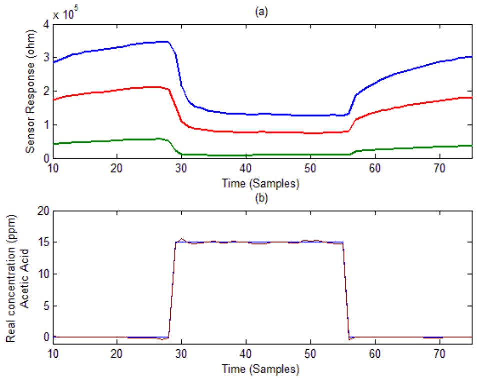
Experimental Setting:

- Two mock Pollutants (Acetic Acid, Ethanol)
- In lab calibration of TinyNose equipped with
- On-Board ANN sw component (NesC).
- Threshold level for Ethanol = 100ppm (/2000ppm)
- p=0.01 probability of positive event

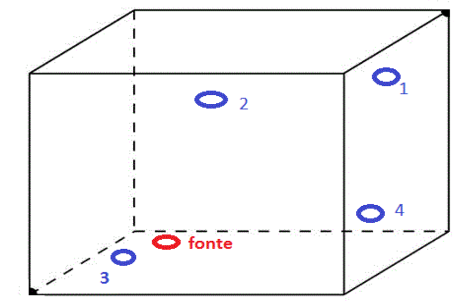
Results:

- Computational footprint tradeoff (2.5mAx25ms)
- 1% False Positive rate
- Extension of lifetime from 47days (1Hz sample f) to 113 days

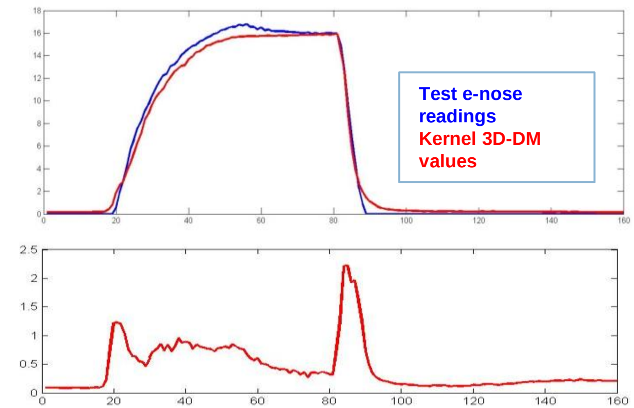
S. De Vito et al. IEEE Sensors Journal, Apr. 2011



Each w-nose was calibrated (in lab) towards the target analytes (in mixture). An ANN component was embedded.

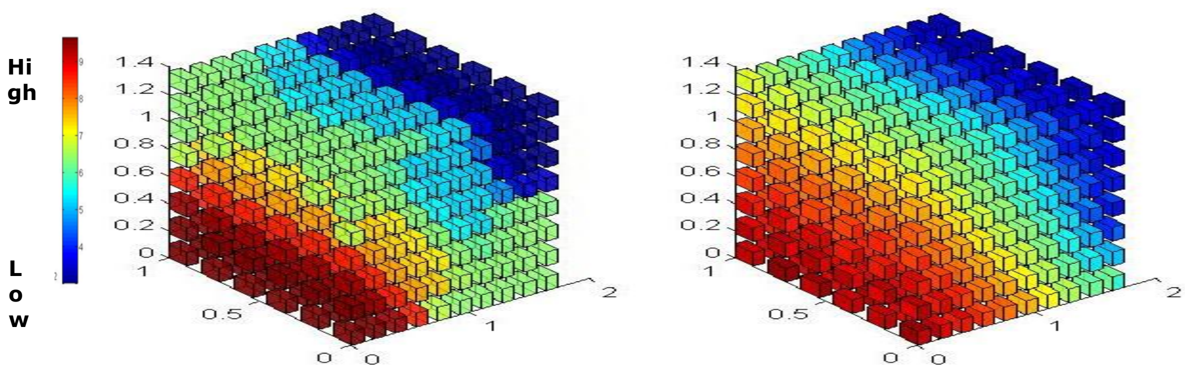


W-noses were deployed in a glass box simulating a 3D ambient. A VOC mixture is let evaporate within the box.



Sensors cross calibrate their Kernel parameters (simulated @ datasink)

3D Reconstruction occurs at Datasink



Istantaneous 3D Ethanol (right) and Acetic Acid (left) concentration images (computed @datasink) using a 4 w-nose deployment in the glass box experimental setup.

Available Facilities

- 3 Climatic Chambers for sensors arrays (2) and sensors nodes (1) characterizations
- Embedded Programming Lab
- Supercomputing (GRID-like) facility
- GIS Models Lab

Suggested R&I Needs for future research

Research directions as R&I NEEDS:

Invest further research energies in:

- **Adaptive (cooperative) Drift Counteraction**
  - Results are not completely satisfactory (still too much time to ignite)
- **Calibration Transfer (Cope with sensor diversity)**
  - You cannot repeat the calibration procedure for each node (Costs)

In order to increase reliability of performance estimation and acceptability of the techniques:

MORE MEASUREMENT DATA!!!

SHARING DATASETS is a KEY NEED

Saverio De Vito, Ph.D.
saverio.devito@enea.it
ENEA / Italy

Thank you

for Your kind attention!

MoniCA
Monitoring
AirQuality Cooperatively

Studio di sistemi *Embedded Pattern Recognition* per *IWCS / Sensor Censoring*

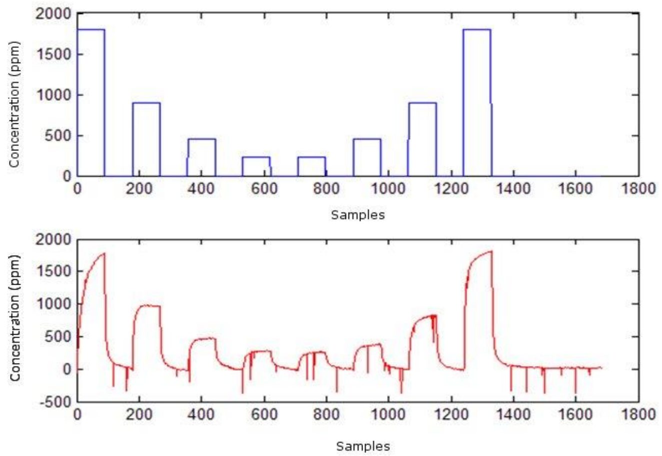


Fig. 3. Acetic acid concentration estimation (red) performed by the FFNN component plotted against true concentration (blue). X axis depict time (samples) while y-axis depicts real and estimated concentrations values.

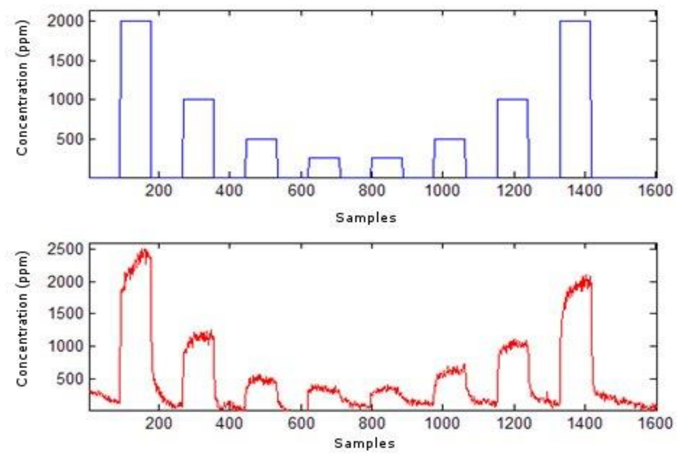


Fig. 4. Ethanol concentration estimation (red) performed by the FFNN component plotted against true concentration (blue). X axis depict time (samples) while y-axis depicts real and estimated concentrations values.

TABLE IV: Memory footprint increase in the Embedded component resulting from the linking of sensor fusion neural component

Algorithm	Bytes in ROM	Bytes in RAM
Basic	20380	574
Basic+NN comp.	27340	910

Modellando come una variabile bernoulliana il risultato della computazione....

X ~ B (1, p)

Si ottiene l' assorbimento medio....

$$I_{cc,mean}^{NN} = I_{RS}\left(\frac{T'_{RS}\,p + T''_{RS}(1-p)}{T}\right) + I_{RW}\frac{T_{RW}}{T}p + $$

$$+ I''_{A}\frac{T_A}{T}(1-p) + I'_{A}\frac{T_A}{T}p + I_T\frac{T_T}{T}p + I_C\frac{T_C}{T}.$$

Con una p=0.01 (alta) si passa da 41 giorni a 113 gironi di expected life per il singolo nodo in conf. Stella.

AIS: Principi…

- **Training: Evolvere un set (compatto) di istanze (codebook) in grado di fungere da istanze rappresentative**

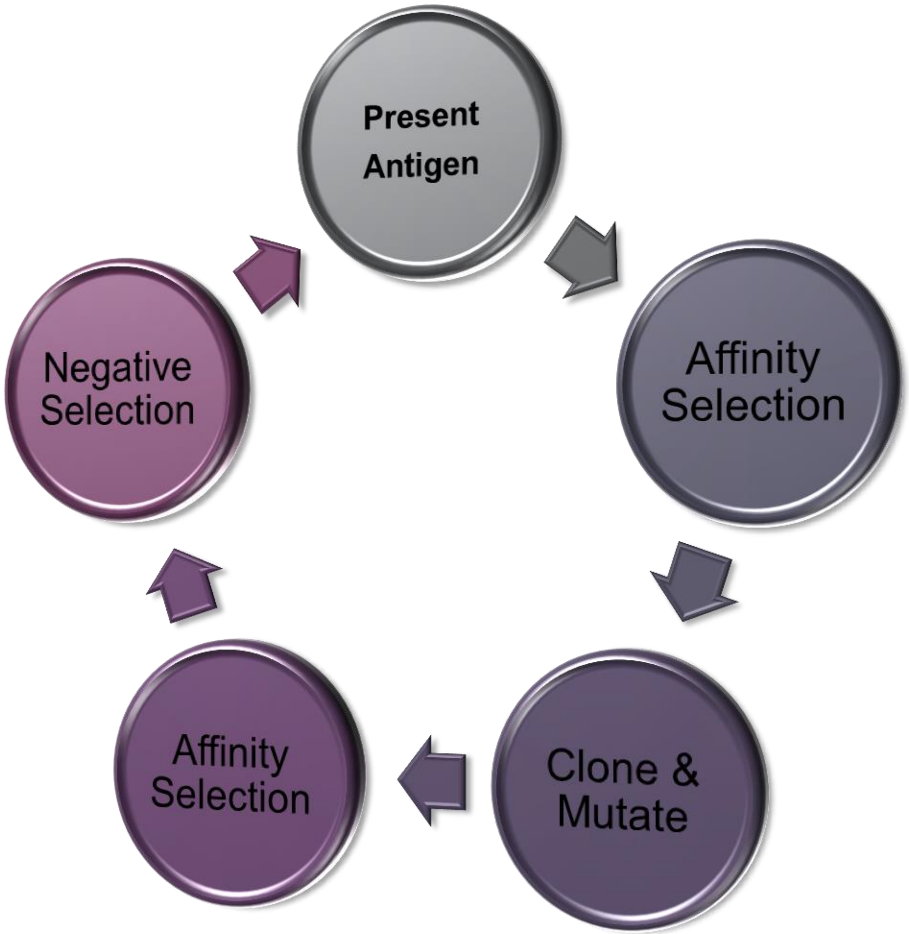
- **Classificare i campioni “unseen” utilizzando il codebook con un algoritmo KNN**

- **Key Factors: How to evolve the codebook?**
  - **Mutation (Creare generazioni di istanze mutate)**
  - **Negative selection (Selezionare x diversità)**
  - **Clonal Expansion (Clonare le istanze “fit”)**

AIS: How they works

- Instance Based Machines
- Antigens and Antibodies are points in the dataset feature space
- Shape space translate in Euclidean Space (though other distance measure have been tested)
- Antigens-Antibody Affinity modeled with distance.

Clone, Mutate, Select



1. Present each Antigen (training instances) to a preliminary antibodies codebook (memory cell set)
2. Select candidates antibodies by affinity (Maximize antigen affinity: low distance)
3. Clone and mutate candidates
4. Select best candidate
5. Prune memory cell set by looking at difference with antibodies in the memory cell set (Minimize self affinity: negative selection)

Architectural Comparison with BPNN

- Classification Mechanism:
  - Hyperplanes vs Nearest Neighbour
- Structure:
  - Neurons (Weights) vs Memory Cells
- Internal Knowledge Representation
  - Weight vs Instances
  - Possible use in EDA of AISs