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

COST Action TD1105

Focus Group Meeting on

Data Analysis of Aveiro Air Quality Sensors Intercomparison

WHO Collaborating Centre (CC) for Air Quality Management and Air Pollution Control - Federal Environment Agency (FEA)

Berlin, Germany, 17 April 2015

Focus group goals and Multivariate calibration

Saverio De Vito

MC Substitute, WG Member, SIG Member

ENEA/ Italy

COST is supported by the EU Framework Programme

ESF provides the COST Office through a European Commission contract

Scientific context and objectives

General Objectives: What we can achieve?

-Performance based comparison of sensor technologies and family across different multisensor devices

-Performance based comparison of calibration techniques across different multisensor devices

-Holistic assessment of technology readiness across different multisensor devices (wrt DQO)

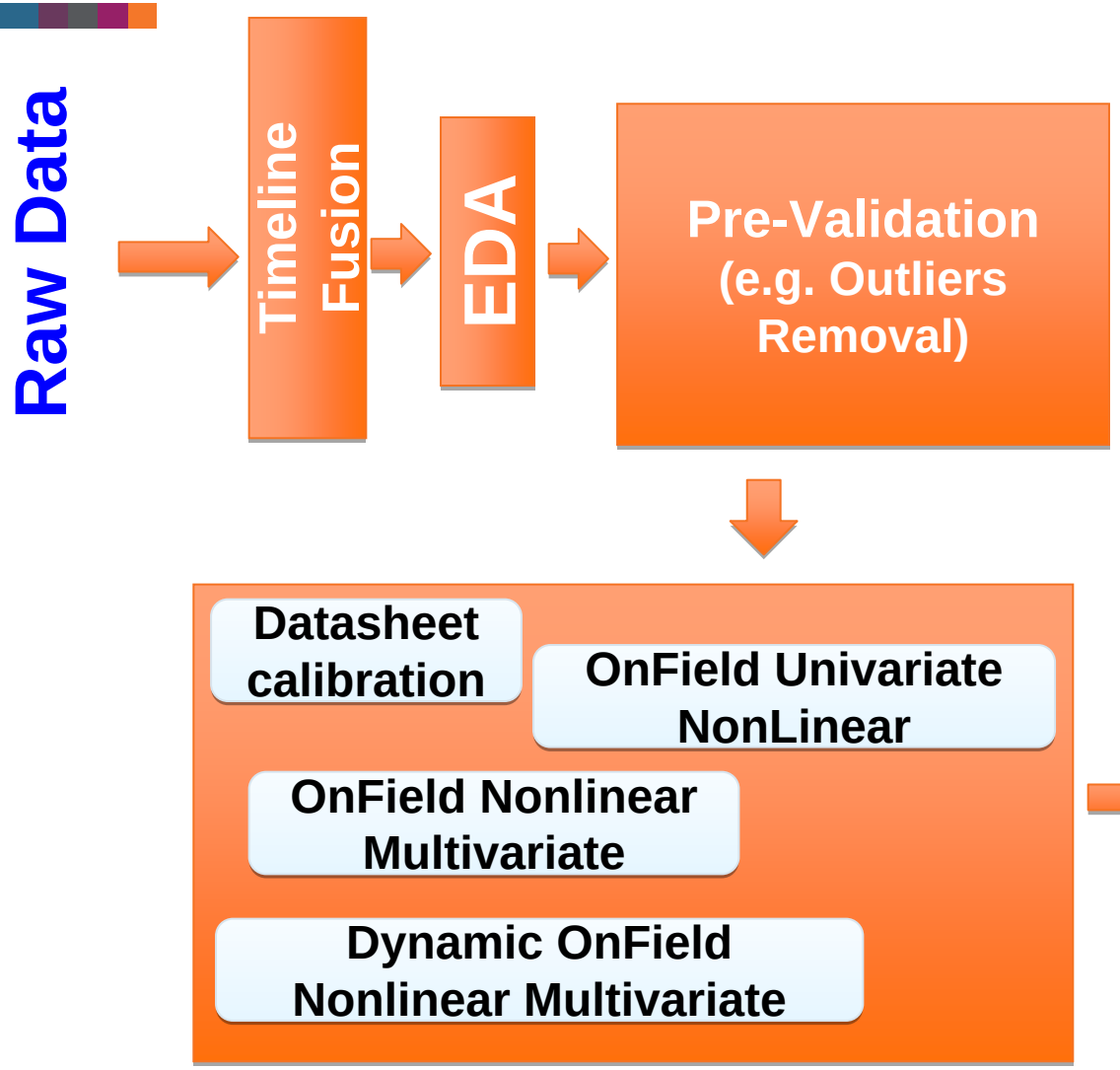
Scientific context and objectives

These assessments are long sought by the scientific and technical community since they can only be achieved with:

A dataset encompassing several etherogeneous multisensor devices measured together in-field across several days.

The Aveiro dataset is a significant value for the EuNetAir Community!

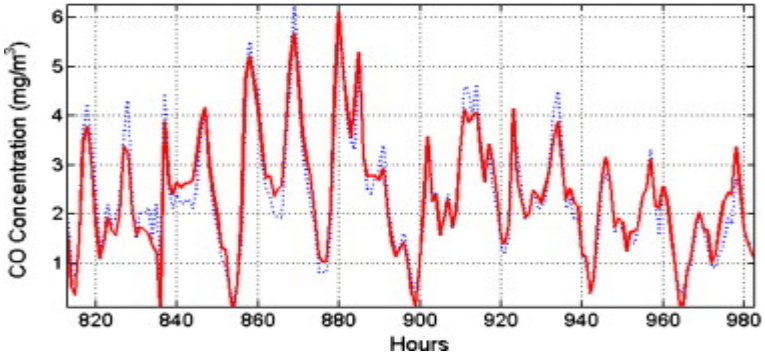
Proposed data analysis roadmap:



Raw Data: Any possible sensors line combination!

Recent achievements

Several works by our group indicate that machine learning based multivariate non linear calibration with on field data is a feasible way to reduce sensor cross sensitivities and stability issues with a limited number of training data.



From 7 days ….. **(De Vito et al, S&A,b 2009)** to

1 day of training data **(De Vito et al., IEEE Sensors, 2012)**

with target instantaneous MRE in the 20%

According to Spinelle et al. from JRC, this technique emerge as the best performing one:

EU Directive DQO Reached!

Input	Intercept	Slope	R^2	RMSE
Raw data	-2.9	1.04	0.95	64.0
Standardized data	-3.9	1.07	0.96	64.8
Calibrated data (MLR)	1.5	0.99	0.95	60.7

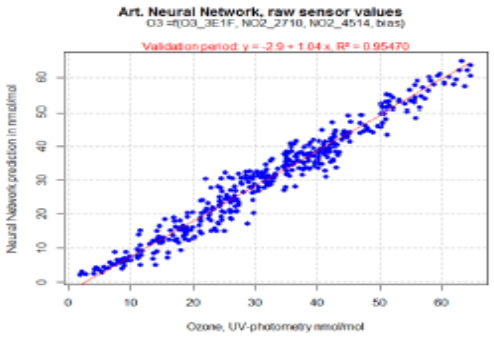


Fig. 3. Scatterplot of the calibrated sensor data using the ANN based on raw sensor's data against reference measurements.

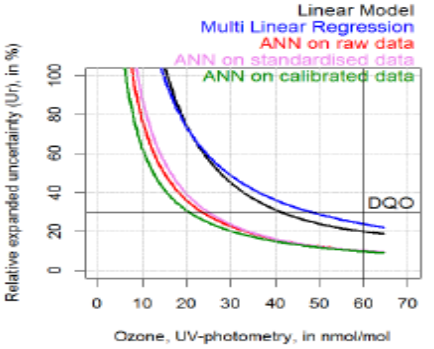
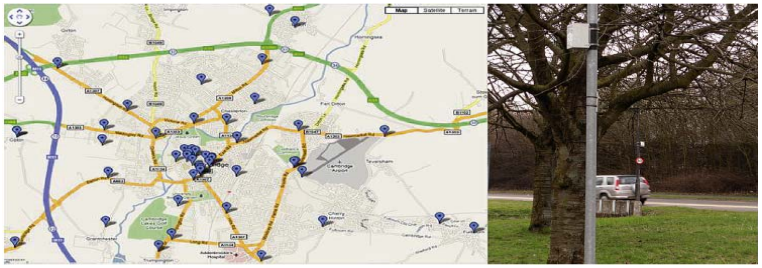
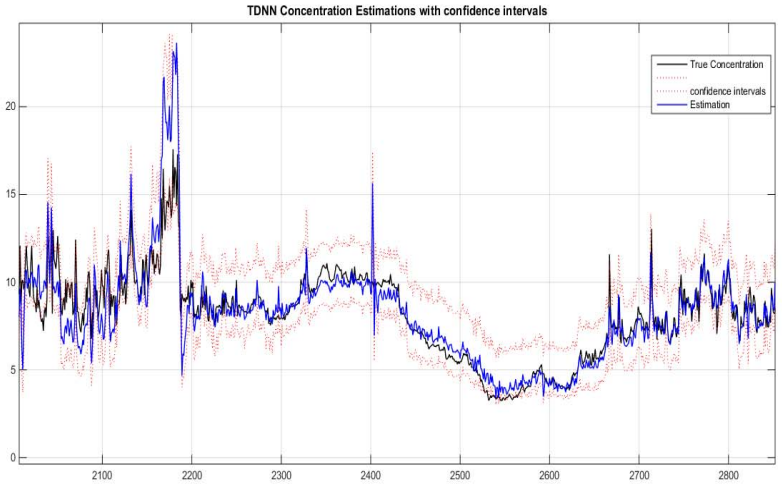


Fig. 4. Relative expanded uncertainty of the estimated values versus reference data as a function of the level of O_3 for the five calibration models.

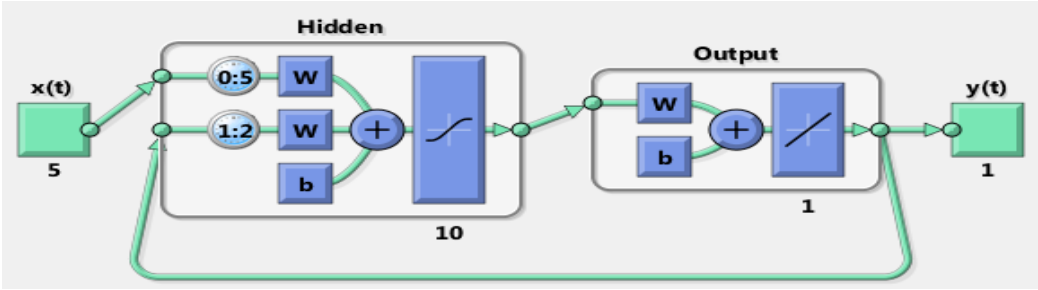
Recent achievements

Recently we started tackling dynamic issues for on-the field deployed, street level or mobile deployments:





ECSensors with Ts=20sec
Inputs: NO,NO_2,O_3,RH,T sensor responses
Output: NO_2 concentration
Neural Auto regressive with exogenous inputs



1 week training test , 1 week validation test , 3 weeks test set.
Hyper parameter search space HN=[3,5,10];
30 repetitions for each HN value, best avg reported.

Static FFNN	TDNN	NARX
1.74 (0.20) ppb	**1.48 (0.13) ppb**	1.50 (0.13) ppb

* In cooperation with CAS-Dept. Of Chemistry- University of Cambridge

DESCRIPTION of Sensors Database to be Used

- The data base structure:
  - Multiple XLS or csv tables (one x each MSD) with sensor data
  - Multisensor device descriptions in terms of adopted sensors and their technology (further details on multisensor device technology if possible)
- Table structure:
  - Raw sensor data
  - Datasheet calibration (if any)
  - Target gases groundtruth as recorded by conventional analyzers
  - Environmental data RH,T, WS (maybe WD)
  - Timing of the measurements
- Information on groundtruth analyzer (sensitivity, MSRE, etc.)
- Target Gases: All available

CONCLUSIONS

- The final goal of your approach for sensors database:
  - Performance based calibration methodology comparisons and technology readiness evaluation
  - Focus on
    - Near-Real time concentration assesmments
    - Nonlinear (dynamic) multivariate calibration (from machine learning field)

- Highlight the *expected achievements* and the open problems
  - Performance goal: less than 20% MRE on instantaneous measurements
  - Reduction of cross sensitivity, non linearity, limited responsiveness issues
  - Calibration performance over multiple instance of the same sensor array (?)
  - Calibration stability over time (not possible on long term due to db limitations)