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

COST Action TD1105

Metrics for the evaluation of AQ sensor performance

Kostas Karatzas

Informatics Applications and Systems – Environmental Informatics Research Group

Dept. of Mechanical Engineering

Aristotle University of Thessaloniki, Greece

kkara@eng.auth.gr, http://isag.meng.auth.gr

COST is supported
by the EU Framework Programme

ESF provides the COST Office
through a European Commission contract

Environmental time series analysis

- Error identification
- Comparison with reference data

Error identification

- Systematic (like time drift)
  - Correct with the aid of reference events
  - Aggregate
- Random
- Semantic

The goal

- Comparison of two time series in order to estimate their resemblance
  - TS1: sensor data
  - TS2: reference method data
  - Resemblance: distance from being identical
    - Distance=quantifiable criteria

Valid for the comparison of any time series (sensor data, model results, other)

The metrics

- Descriptive (basic) statistics
  - Min, Max, Median, Mean, Std, Skewness, Kurtosis

Comparison statistics

$$MBE = \overline{M} - \overline{RM}$$

Mean systematic difference; magnitude of differences between sensors estimation and reference values averaged over the whole sampling period

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|M_i - RM_i\right|$$

indicates the average of the magnitude of the errors; is sensitive to outlier errors.
Can be normalized with the standard deviation of the observations from the reference instrument.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(M_i - RM_i\right)^2}$$

indicates the magnitude of the error and retains the variable's unit; varies with the std of RM; is sensitive to extreme values and to outliers; tends to vary as a function of the standard deviation of the RM

$$Corr.Coef = r = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(M_i - \overline{M}\right)\left(M_i - \overline{RM}\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(M_i - \overline{M}\right)^2 \frac{1}{n}\sum_{i=1}^{n}\left(RM_i - \overline{RM}\right)^2}}$$

measures the strength and the direction of a linear relationship between two variables

Comparison statistics

$$NMSE = \frac{\sum_{i=1}^{n}(M_i - RM_i)^2}{\sum_{i=1}^{n}(M_i - \overline{M})^2} = \frac{\mu^2_{M-RM}}{\mu^2_{M}}$$

an estimator of the overall deviations between reference and sensor measurements; is sensitive to extreme values

$$FB = \frac{\mu^2_M - \mu^2_{RM}}{\frac{1}{2}(\mu^2_M + \mu^2_{RM})}$$

a measure of the agreement between the mean measured concentrations against the reference measurements; perfect agreement: FB=NMSE=0

$$FOEX = 100 \times \left[\frac{N(M_i > RM_i)}{N_{total}} - \frac{1}{2}\right]$$

a measure of the over or under estimation of studied measurements against reference data. The best condition would be that FOEX = 0, unless, ideally, all measurements are equal to reference measurements, thus FOEX would be equal to -50.

$$CRMSE = \sqrt{\frac{1}{n}\sum_{1}^{N}[(M_i - \overline{M}) - (RM_i - \overline{RM})]^2}$$

used for the quadratic decomposition of RMSE as the sum of Mean Bias Error and Centered Root Mean Error; is an indicator of the sensor random error. Can be normalised with the standard deviation of RM.

And there are more candidates…

$$d = 1 - \frac{\sum_{i=1}^{n} \left| M_i - RM_i \right|^2}{\sum_{i=1}^{n} \left(\left| M_i - \overline{RM} \right| + \left| RM_i - \overline{RM} \right| \right)^2}$$

$$d1 = 1 - \frac{\sum_{i=1}^{n} \left| M_i - RM_i \right|}{\sum_{i=1}^{n} \left(\left| M_i - \overline{RM} \right| + \left| RM_i - \overline{RM} \right| \right)}$$

Willmott's index of agreement

$$M = (2/\pi)\sin^{-1}\left\{ 1 - \frac{MSE}{Var_M + Var_{RM} + (\overline{M} - \overline{RM})^2} \right\}$$

Watterson's (1996) measure of agreement (M

$$\Re = 1 - \frac{\delta}{\mu}$$

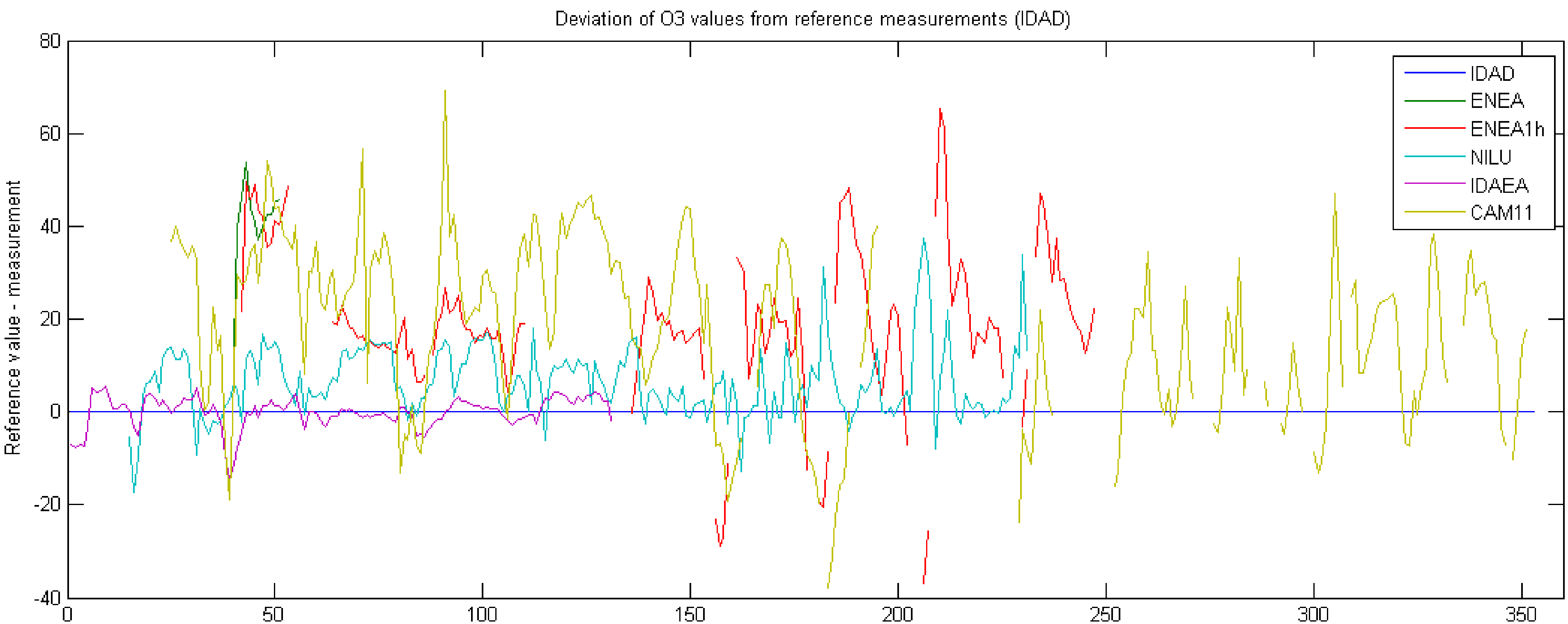
$$\delta = MAE$$

$$\mu = n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| M_j - RM_i \right|$$

Mielke & Berry (2007)

Kyriakidis I., Kukkonen J., Karatzas K., Papadourakis G.and Ware A. (2015), New Statistical Indices for Evaluating Model Forecasting Performance. Proceedings of the 9th International Conference in New Horizons in Industry, Business and Education, Skiathos Island, Greece. Publisher: Technological Educational Institute of Crete, ISBN: 978-960-99889-9-5, pp. 104-118 (available at http://nhibe2015.vs-net.eu/proceedings/papers/3_10_[P]0065.pdf).

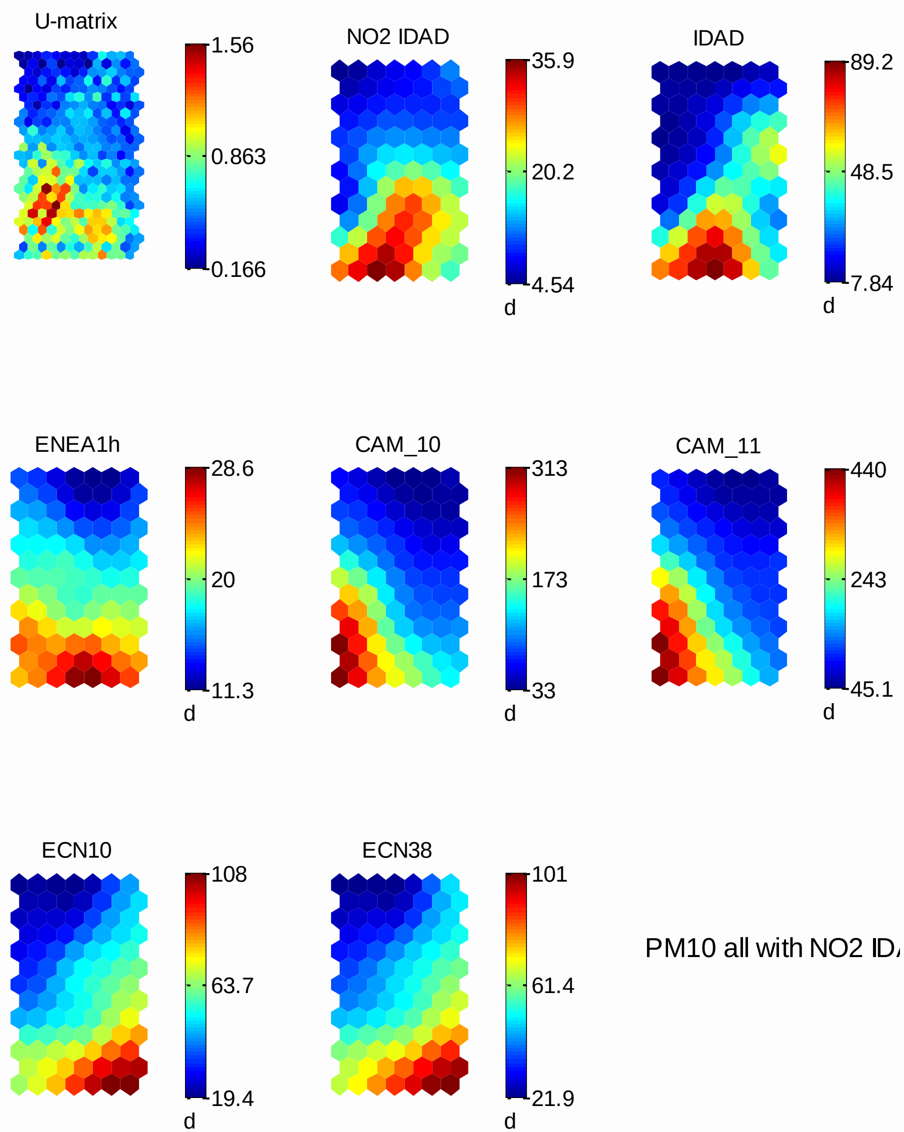
So how can we use them?



Sensor node	MBE	r^2	$CRMSE/\sigma_0$	MBE/σ_0	NMSE	FB	FOEX	MAE
ENEA	19.2	0.13	1.60	1.90	2.52	0.67	41.61	22.12
NILU	6.5	0.77	0.84	0.72	3.86	0.33	32.95	7.66
CAM_11	15.7	0.14	1.95	1.68	2.74	0.71	30.23	21.50
IDAEA/AQMesh	0.0	0.70	-0.55	-0.01	4.25	0.19	7.25	2.40
ISAG	356.1	0.12	17.49	33.24	1.37	1.82	50	360.12

Informatics Systems & Applications – Environmental Informatics Group

Data analytics: data insights with CI

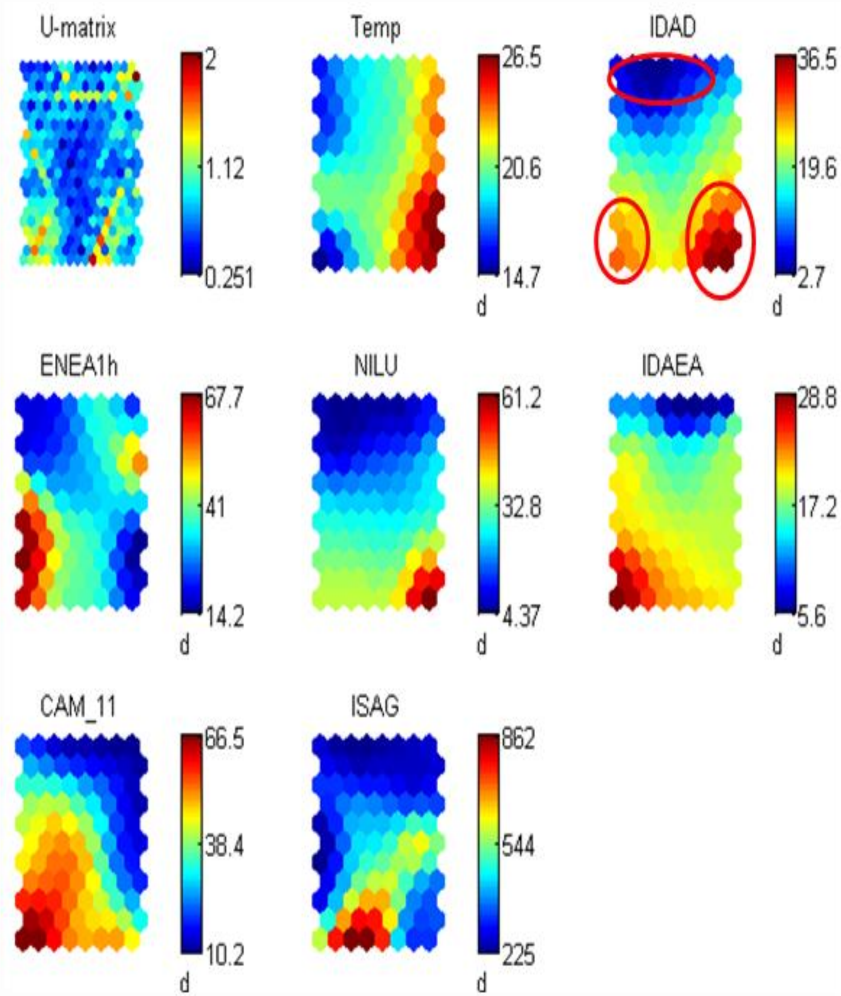


PM10:

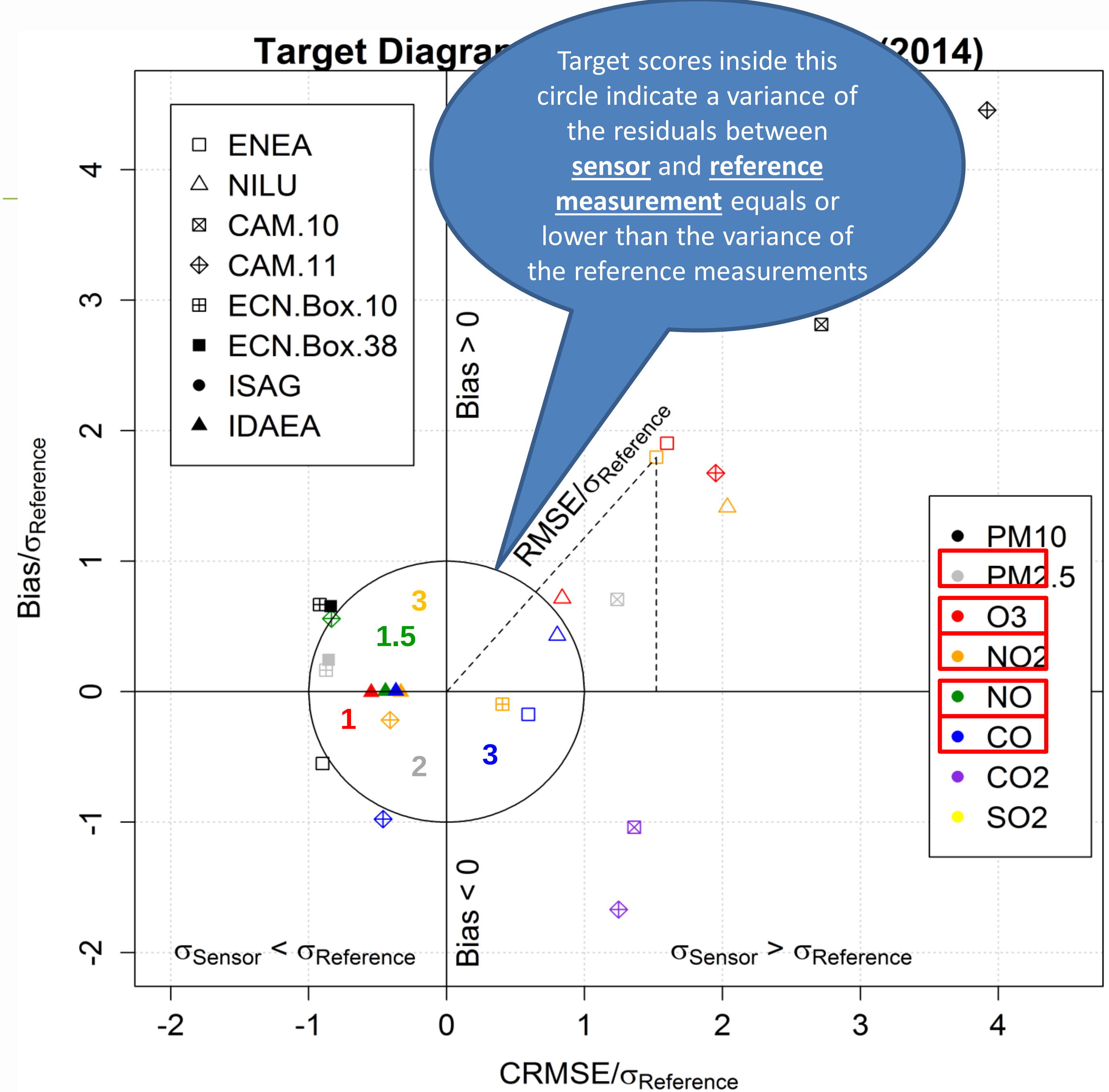
-relationship with NO2

-pattern resemblance with ENEA

Data analytics: data insights with CI



- low Ozone conditions (medium to low temp)
- High Ozone, high temp
- Med. to high Ozone, low temp (O3 transportation??)



Target Diagra
(2014)
Target scores inside this circle indicate a variance of the residuals between **sensor** and **reference measurement** equals or lower than the variance of the reference measurements
ENEA
NILU
CAM.10
CAM.11
ECN.Box.10
ECN.Box.38
ISAG
IDAEA
PM10
PM2.5
O3
NO2
NO
CO
CO2
SO2
Bias > 0
Bias < 0
RMSE/σReference
Bias/σReference
CRMSE/σReference
σSensor < σReference
σSensor > σReference
3
1.5
1
2
3
-2
-1
0
1
2
3
4

Conclusion

- A variety of metrics for data description should be tested
  - Descriptive statistics tell part of the story
- Advanced data analytics based on CI to be also used
  - Unsupervised learning methods are advantageous
- A comparison method applied for all?
  - Yes, if carefully selected and thoroughly evaluated

ACKNOWLEDGEMENTS

Thank you!