Estimation of ambient air levels of regulated heavy metals by means of Partial Least Squares Regression (PLSR)

Germán Santos
Ignacio Fernández-Olmo

Department of Chemical Engineering and Inorganic Chemistry
University of Cantabria (Spain)
Development of Chemical Processes and Pollutants Control Research Group (DePRO)

Sosprocan DePRO
Desarrollo de Procesos Químicos y Control de Contaminantes

May 9th, 2013
Madrid, Spain

Contents

1. Introduction

1.1. Environmental regulations

- Air Quality Framework Directive[1]
- Prevent or reduce harmful effects to human health and the environment
  - Ambient air quality objectives
  - Air quality assessment (legal duty)
  - Air quality maintenance/improvement

Assessment methods

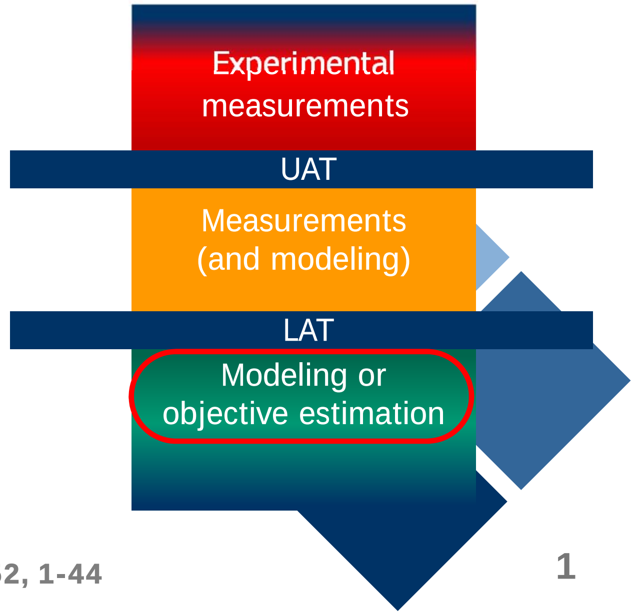


Table 1. UE quality objectives and evaluation thresholds for regulated metals

Pollutant	TV[a] (ng m^{-3})	UAT[b]	LAT[b]	Directive
Pb	**500**	**60**	**40**	**2008/50/EC**
As	**6**	**60**	**40**	
Cd	**5**	**60**	**40**	**2004/107/EC**
Ni	**20**	**70**	**50**	

TV: Target Value; UAT: Upper Assessment Threshold; LAT: Lower Assessment Threshold

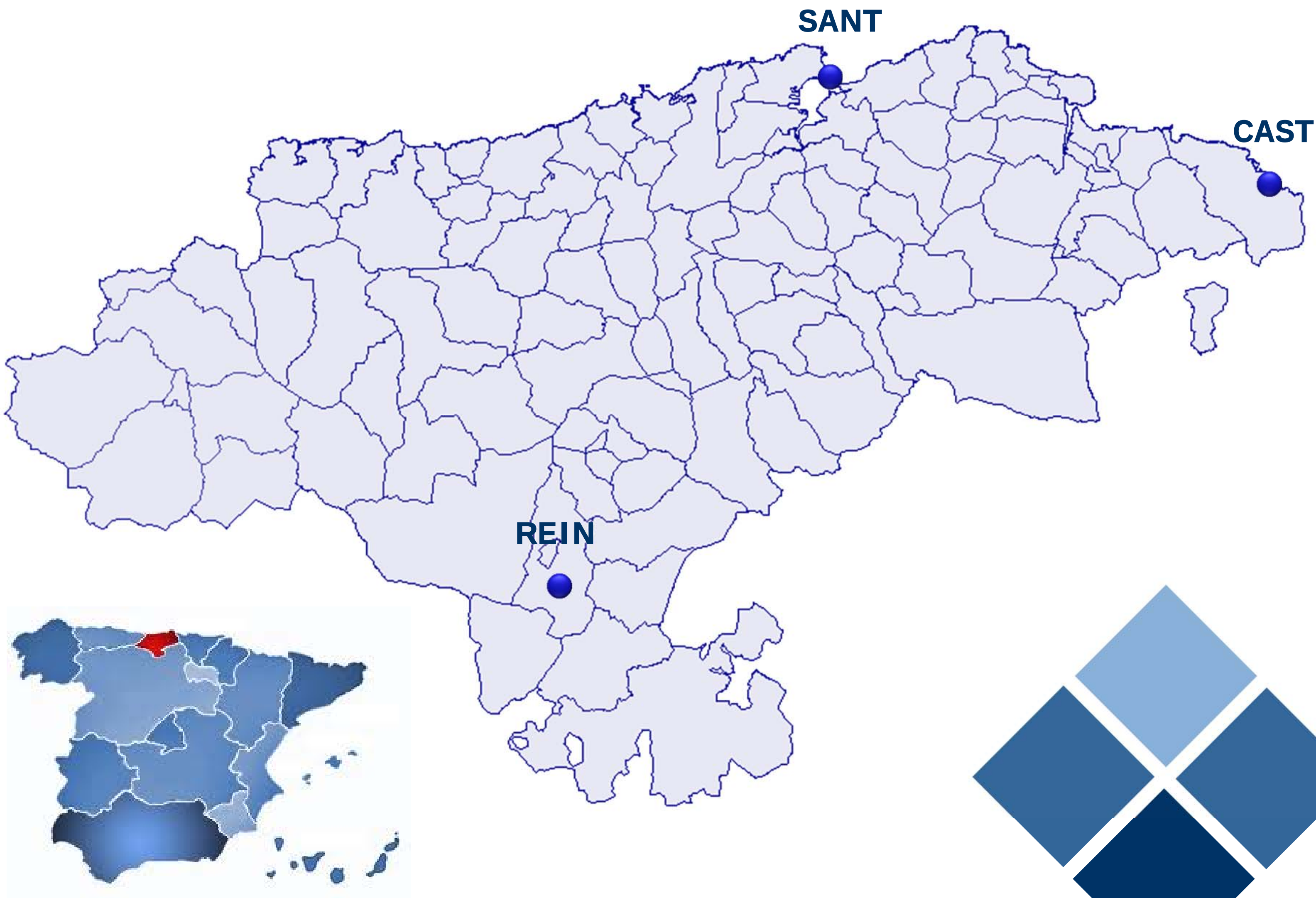
[a] For the total content in the PM_{10} fraction averaged over a calendar year

[b] Percent of the target value

[1] European Commission (EC), *Off. J. Eur. Communities: Legis.*, 2008, 152, 1-44

1. Introduction

1.2. Area of study



1. Introduction

1.2. Area of study

1.3. Air quality assessment in Cantabria

Objective estimation?

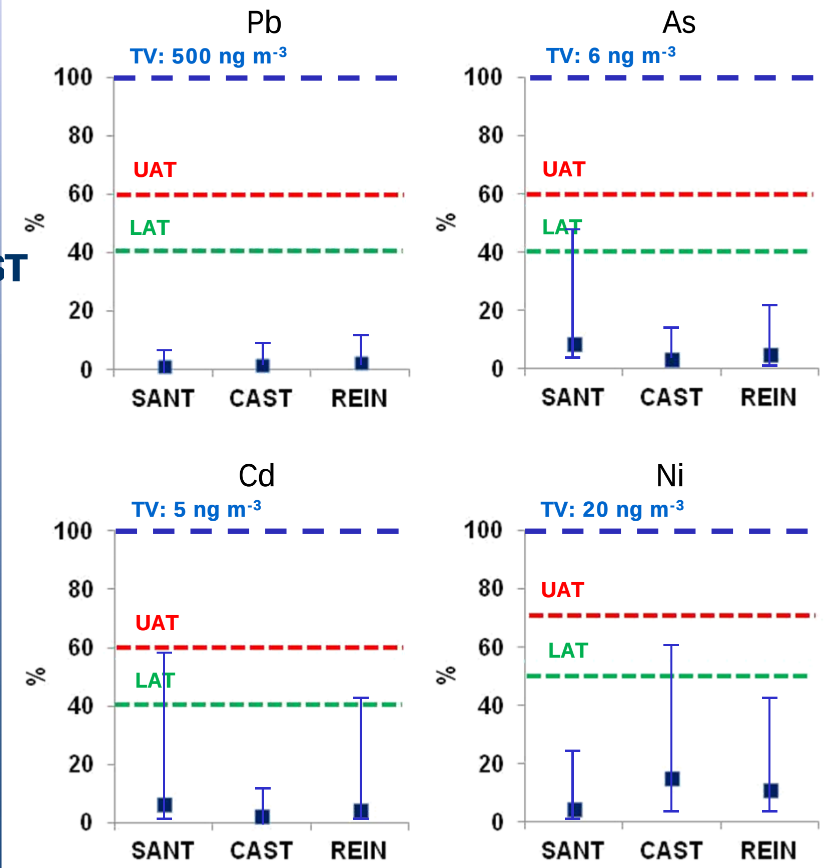


Figure 1. 2008 annual mean of regulated metals in PM_{10} expressed as percent of their respective target values [2]

[2] A. Arruti, I. Fernández-Olmo and A. Irabien, *J. Environ. Monit.*, 2011,13(7), 1991-2000

1.4. Aim of this work

Estimate the annual levels of the EU regulated metals in airborne PM_{10} in urban areas in the Cantabria Region (Northern Spain)

- Development of statistical models based on Partial Least Squares Regression (PLSR)
- Comparison between the estimated metal levels using PLSR and Multiple Linear Regression (MLR) and Principal Components Regression (PCR), from previous works

~~Forecasting~~

Estimation

2. Development of PLSR models

2.1.Why PLS?

Model a target variable (response) when there are a large number of predictor variables

MLR creates a linear combination of the predictors that best correlates with the response

PCR creates linear combinations (components) of the predictors with large variance, reducing correlations, without using the response values. Then uses those combinations in predicting the target variable instead of the original predictors

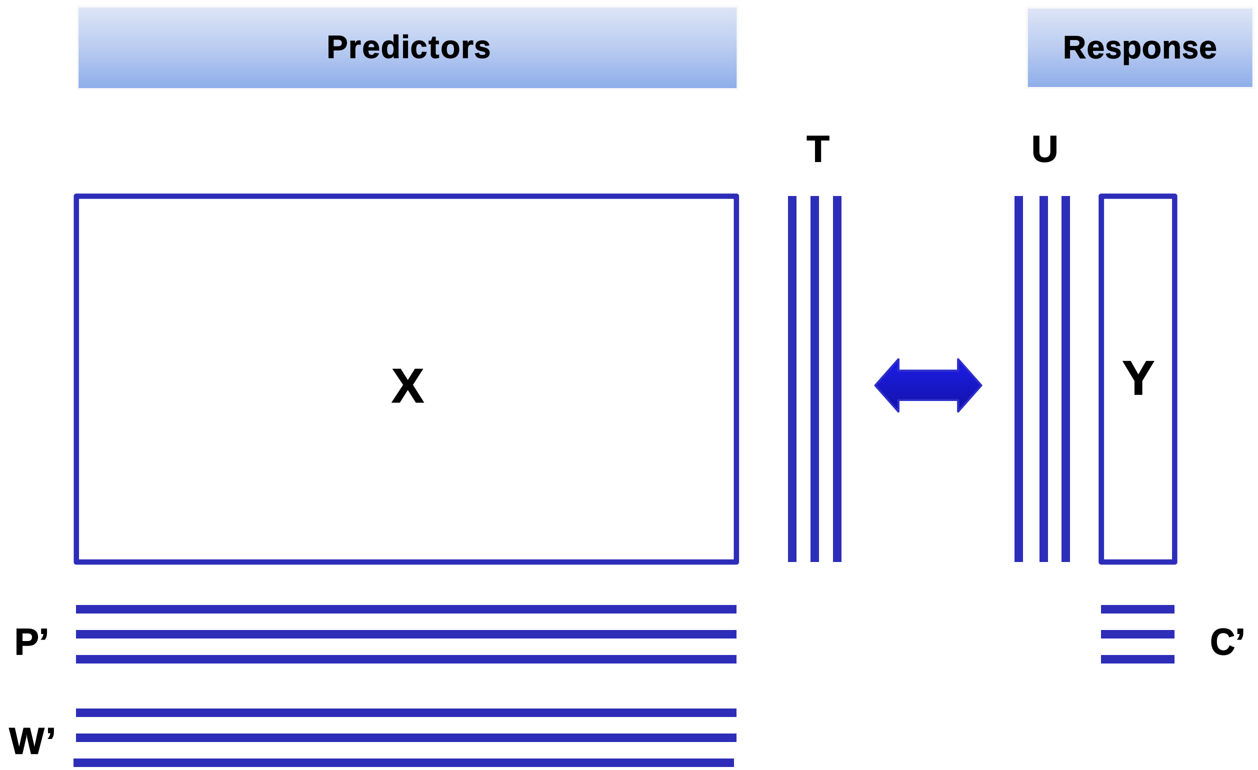
PLSR creates new predictor variables, latent variables (LVs), as linear combinations of the original predictors, as PCR does. The difference is on how the components are computed

PCR weights are calculated from the covariance matrix of the predictors, while PLSR weights reflect the covariance structure between predictors and response

PLSR combines information about the variances of both the predictors and the responses, while also considering the correlations among them

2. Development of PLSR models

2.2. PLSR fundamentals



T X-scores
U Y-scores
P’ X-loadings
C’ Y-loadings
W’ weights

$$X = T\,P' + E$$
$$T = X\,W^*$$

$$Y = U\,C' + G$$
$$Y = T\,C' + F$$

$$Y = X\,W^*\,C' + F = X\,B + F$$

Figure 2. Matrix structure of PLSR

2. Development of PLSR models

2.3. Input data

Training dataset

PREDICTORS

RESPONSES

Nominal SE SD WE | T RH WD PP | SO_2 NO_x O_3 PM_{10} | Pb As Cd Ni

SE	S
SD	S
WE	W

Continu

Meteoro

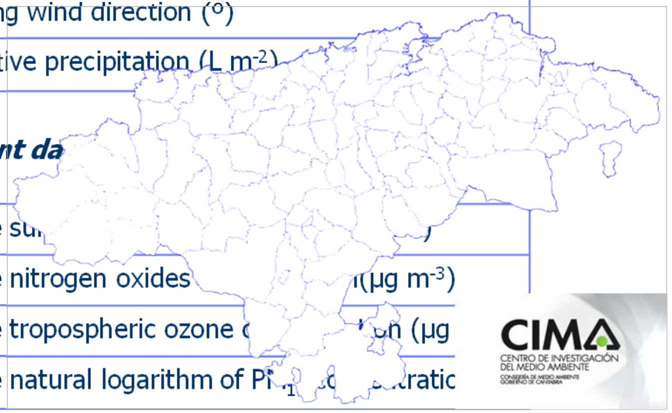
T	A
RH	Average relative humidity (%)
WD	Prevailing wind direction (º)
PP	Cumulative precipitation (L m^{-2})

Major pollutant da

SO_2	Average su
NO_x	Average nitrogen oxides (µg m^{-3})
O_3	Average tropospheric ozone (µg
PM_{10}	Average natural logarithm of PM$_{10}$ concentratio

Pb	)
As	Ar m^{-3})
Cd	Ca m^{-3})
Ni	Ni ^{3})

Regional Air Quality Monitoring Network



CIMA CENTRO DE INVESTIGACIÓN DEL MEDIO AMBIENTE

Pre-Proces	**Auto-scaling**
. LVs	**Cross-Validation**
Software	**PLS Too**

PLS_Toolbox

2.4. Model performance criteria

Quality Indicators

Correlation coefficient (1)

$$r = \left[\frac{\sum_{i=1}^{n}(C_{O,i} - \overline{C_O})(C_{E,i} - \overline{C_E})}{\sqrt{\sigma_O \sigma_E}}\right]$$

Fractional Bias (0)

$$FB = \frac{\overline{C_O} - \overline{C_E}}{0.5\,(\overline{C_O} + \overline{C_E})}$$

Root Mean Square Error (0)

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(C_{O,i} - C_{E,i})^2}$$

Normalized Mean Square Error (0)

$$NMSE = \frac{\overline{(C_O - C_E)^2}}{\overline{C_O}\,\overline{C_E}}$$

Fractional Variance (0)

$$FV = 2\,\frac{\sigma_O - \sigma_E}{\sigma_O - \sigma_E}$$

Uncertainty according to EU Directives[a]

$$RME = \max(|C_{O,p} - C_{E,p}|)/C_{O,p}$$

Relative Maximum Error without timing

$$RDE = |C_{O,LV} - C_{E,LV}|/LV$$

Relative Directive Error

[a] EU Uncertainty requirements for objective estimations: RME and RDE < 100%

3. Performance of PLSR models

Table 2. Performance indexes for the estimations

Metal	Annual mean ($ng\ m^{-3}$)		EU uncertainty		Performance indexes				
	Observed	Estimated	RME (%)	RDE (%)	r	FB	RMSE	FV	NMSE
CAST									
Pb	8.0	8.0	19	1.5	0.9	0.0	3.2	0.15	0.2
As	0.2	0.2	49	3.9	0.7	0.0	0.1	0.77	0.8
Ni	3.0	3.0	34	22	0.8	0.0	1.7	0.48	0.3
Cd	0.1	0.1	34	3.9	0.8	0.0	0.1	0.64	0.7
REIN									
Pb	11.2	11.2	18	1.0	0.9	0.0	4.2	10	0.1
As	0.3	0.3	55	3.8	0.8	0.0	0.2	0.36	0.2
Ni	2.0	2.0	95	6.9	0.9	0.0	0.7	0.28	0.1
Cd	0.2	0.2	22	10.0	0.9	0.0	0.2	0.22	1.2
SANT									
Pb	6.4	6.4	60	3.8	0.6	0.0	5.7	1.06	0.8
As	0.8	0.8	32	79	0.8	0.0	1.2	0.38	2.1
Ni	0.9	0.9	68	14	0.5	0.0	0.7	1.16	0.6
Cd	0.3	0.3	70	41	0.5	0.0	0.4	1.32	2.8

Comparison of multivariate regression techniques

Table 3. Performance indexes for the estimations

Metal	Technique	Annual mean (ng m^{-3})		Performance indexes					EU uncertainty	
		Observed	Estimated	r	FB	RMSE	FV	NMSE	RME (%)	RDE (%)
CAST										
Pb	MLR	8.0	8.4	0.9	0.0	17.4	0.26	0.1	20	1.5
	PLSR	8.0	8.0	0.9	0.0	3.2	0.15	0.2	19	1.5
As	MLR	0.2	0.2	0.6	0.0	0.0	0.85	0.8	48	4.0
	PLSR	0.2	0.2	0.7	0.0	0.1	0.77	0.8	49	3.9
Ni	MLR	3.0	3.0	0.8	0.0	2.2	0.48	0.3	36	22
	PLSR	3.0	3.0	0.8	0.0	1.7	0.48	0.3	34	22
Cd	MLR	0.1	0.1	0.8	-0.1	0.1	0.63	0.7	29	3.2
	PLSR	0.1	0.1	0.8	0.0	0.1	0.64	0.7	34	3.9
REIN										
Pb	MLR	11.2	11.2	0.9	0.0	[illegible]	[illegible]	[illegible]	16	0.3
	PLSR	11.2	11.2	0.9	0.0	4.2	[illegible]	[illegible]	18	1.0
As	MLR	0.3	0.3	0.8	0.0	0.2	0.42	0.3	35	1.8
	PLSR	0.3	0.3	0.8	0.0	0.2	0.36	0.2	55	3.8
Ni	MLR	2.0	2.0	0.8	0.0	0.8	0.42	0.2	[illegible]	[illegible]
	PLSR	2.0	2.0	0.9	0.0	0.7	0.28	0.1	[illegible]	[illegible]
Cd	MLR	0.2	0.2	0.9	-0.2	0.2	0.37	0.8	[illegible]	[illegible]
	PLSR	0.2	0.2	0.9	0.0	0.2	0.22	1.2	22	10.0
SANT										
Pb	MLR	6.4	6.3	0.6	0.0	5.6	1.07	0.8	48	3.2
	PCR	6.4	6.4	0.5	0.0	6.0	1.27	0.9	59	2.1
	PLSR	6.4	6.4	0.6	0.0	5.7	1.06	0.8	60	3.8
As	MLR	0.8	0.9	0.8	0.11	1.1	0.45	1.8	33	69
	PCR	0.8	0.9	0.6	[illegible]	[illegible]	[illegible]	3.6	67	42
	PLSR	0.8	0.8	0.8	[illegible]	[illegible]	[illegible]	2.1	32	79
Ni	MLR	0.9	0.9	0.5	0.0	0.7	1.24	0.6	59	12
	PCR	0.9	0.9	0.4	0.0	0.7	1.50	0.7	55	12
	PLSR	0.9	0.9	0.5	0.0	0.7	1.16	0.6	68	14
Cd	MLR	0.3	0.2	0.4	0.0	0.4	1.32	2.9	72	42
	PCR	0.3	0.3	0.4	-1.8	4.0	1.82	181	259	151
	PLSR	0.3	0.3	0.5	0.0	0.4	1.32	2.8	70	41

r: 0.6-09

<100%

r: 0.4-08

4. Comparison of multivariate regression techniques

Daily variations

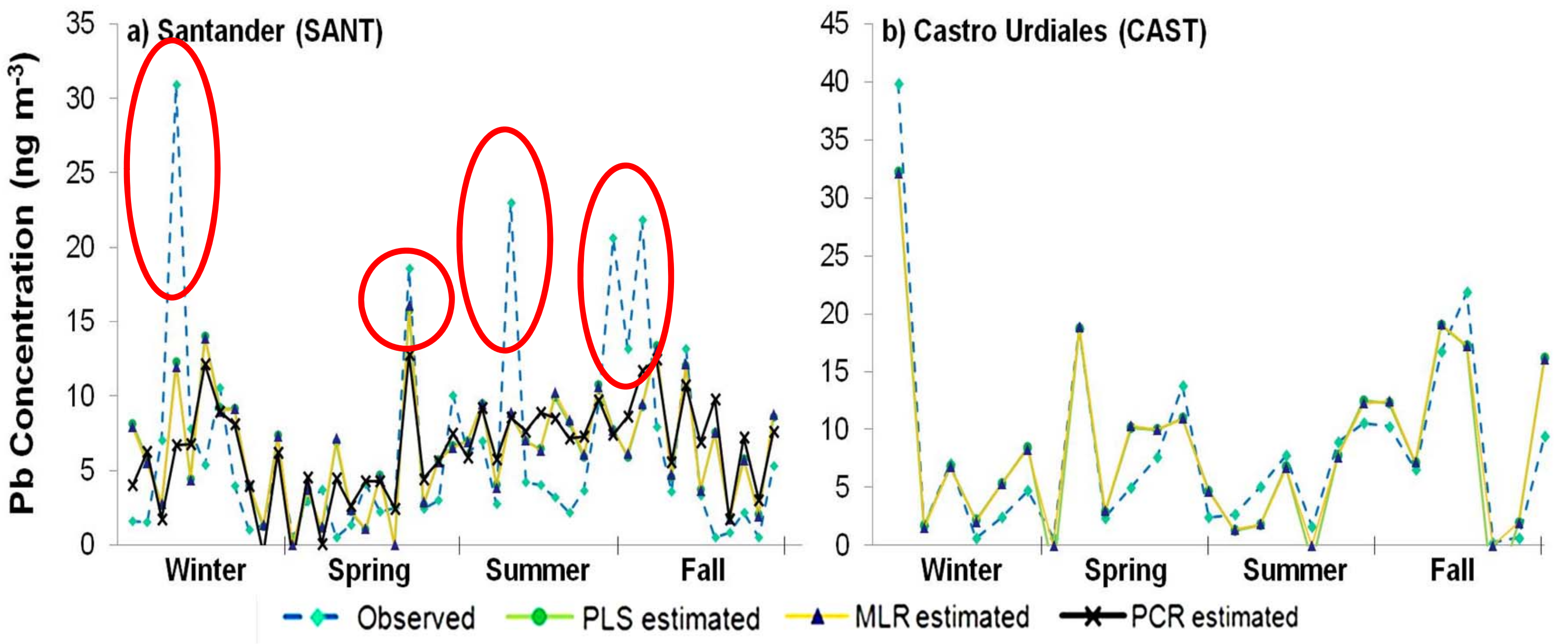


Figure 3. Comparison between observed and MLR, PCR and PLSR estimated daily mean concentration of Pb at the SANT site

Figure 4. Estimated and observed daily mean concentration of Pb at the CAST site

	NMSE	FV
MLR	0.8	1.07
PCR	0.9	1.27
PLSR	0.8	1.06

	NMSE	FV
MLR	0.1	0.26
PLSR	0.2	0.15

5. Conclusions

Model performance

- ✓ PLSR and MLR techniques estimates the regulated metals mean concentrations correctly

Uncertainty requirements

- ✓ PLSR and MLR estimations of the regulated metals fulfill the uncertainty requirements for objective estimations (lower than 100%)

Objective estimation techniques

- ✓ Statistical estimation models based on MLR and PLSR could be employed to assess the air quality at the considered urban areas as an alternative to experimental measurements

Further work

- ✓ Application of more powerful estimation tools (e.g. neural networks)
- ✓ Development of estimations of non-regulated metals with higher concentration levels on ambient air (e.g. Mn or Zn), which will [illegible]nd more strict uncertainty requirements

Thank you for your attention!

Department of Chemical Engineering and Inorganic Chemistry
University of Cantabria (Spain)
Development of Chemical Processes and Pollutants Control Research Group (DePRO)

May 9th, 2013
Madrid, Spain

Department of Chemical Engineering and Inorganic Chemistry

Universidad de Cantabria (SPAIN)

Development of Chemical Processes and Pollutants Control Research Group (DePRO)

Dpt. Ingeniería Química y Química Inorgánica

ETSIIyT, Avda. los Castros s/n

Santander, Cantabria. SPAIN

Tel. +34 942 201597 Fax +34 942 201591

«We know what we have to do to achieve the sustainability, now it's time to do it »