Rijksinstituut voor Volksgezondheid
en Milieu
*Ministerie van Volksgezondheid,
Welzijn en Sport*

National Institute for Public Health and
the Environment
Ministry of Health, Welfare and Sport

PCA (Principal components analysis) to evaluate representativeness of the Dutch monitoring sites

Contents

- Classification of measuring stations
- Use of Principal Components Analysis
- Concluding remarks

Original work:

Evaluation of the representativeness of the Dutch air quality monitoring stations

The National, Amsterdam, Noord-Holland, Rijnmond-area, Limburg and Noord-Brabant networks,

RIVM Report 680704021/2012

P.L. Nguyen, RIVM G. Stefess, RIVM D. de Jonge, GGD Amsterdam A. Snijder, DCMR P.M.J.A. Hermans, Province of Limburg S. van Loon, Province of Noord-Brabant R. Hoogerbrugge, RIVM

Introduction

- There are several monitoring networks in the Netherlands:
  - National (National Institute for Public Health and the Environment)
  - Local environmental agencies
  - Large municipalities

- Stations in the Dutch Networks are classified in several types:
  - Rural
  - (Sub)Urban background
  - Traffic
  - Industry (not a formal classification)

- Quite similar to the classification used in Airbase.

Introduction

- In many cases the type of station is clear.

- However, in some cases questions:
  - Outskirts of city (urban?) versus rural
  - Located near road (traffic?) with a limited amount of traffic
  - Urban location near industries
  - etc.

- Principal Components Analysis is one of the tools to help determine characteristics of measuring stations.

- Observed similar characteristics can be used to classify stations.

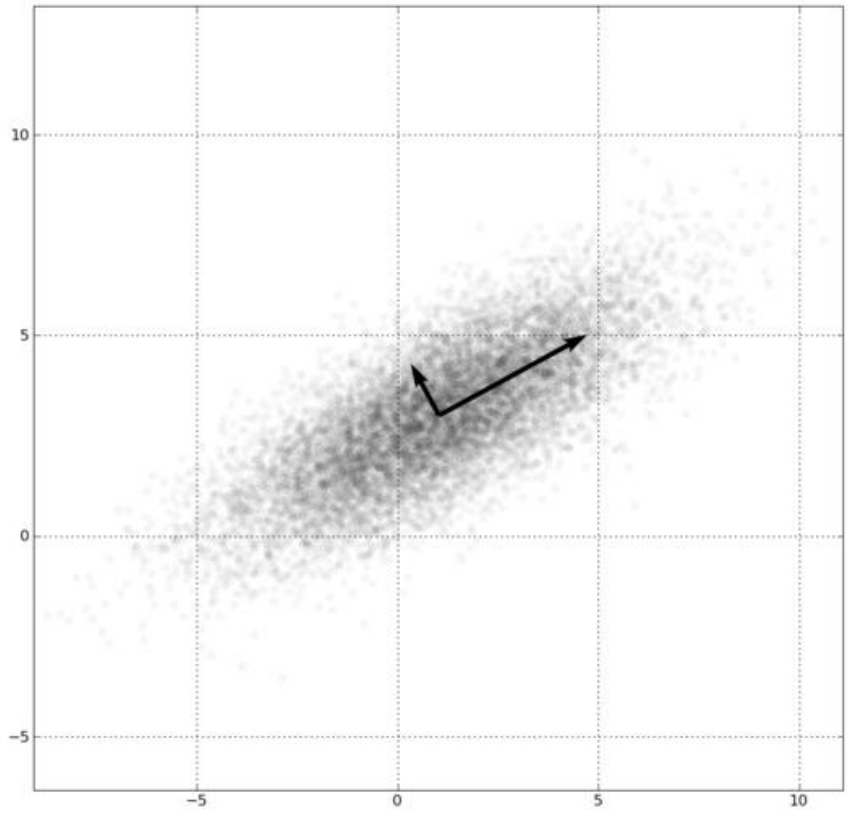
Principal component analysis

- Principal component analysis (PCA) is a statistical procedure that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components.

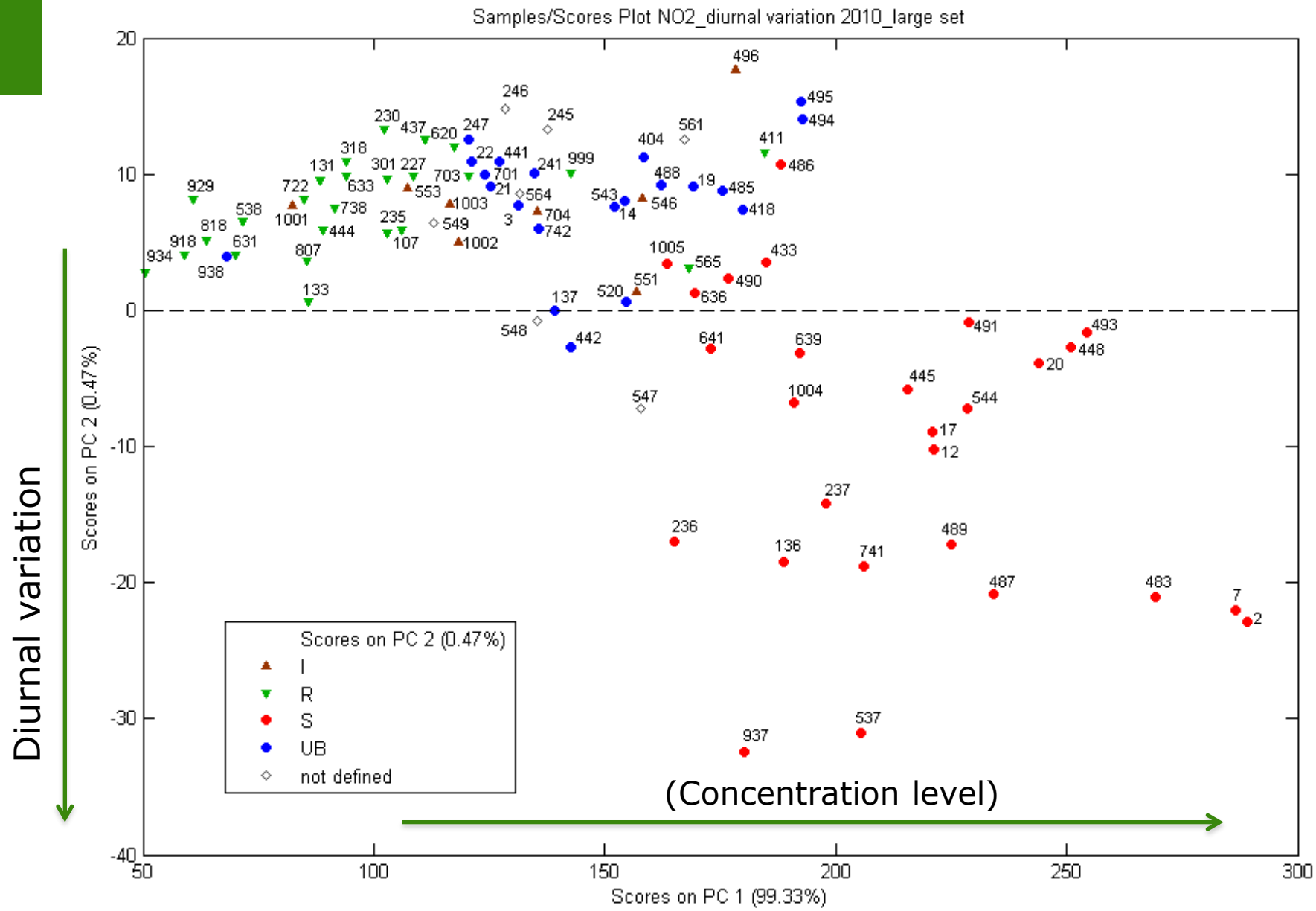
- This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal to (i.e., uncorrelated with) the preceding components.

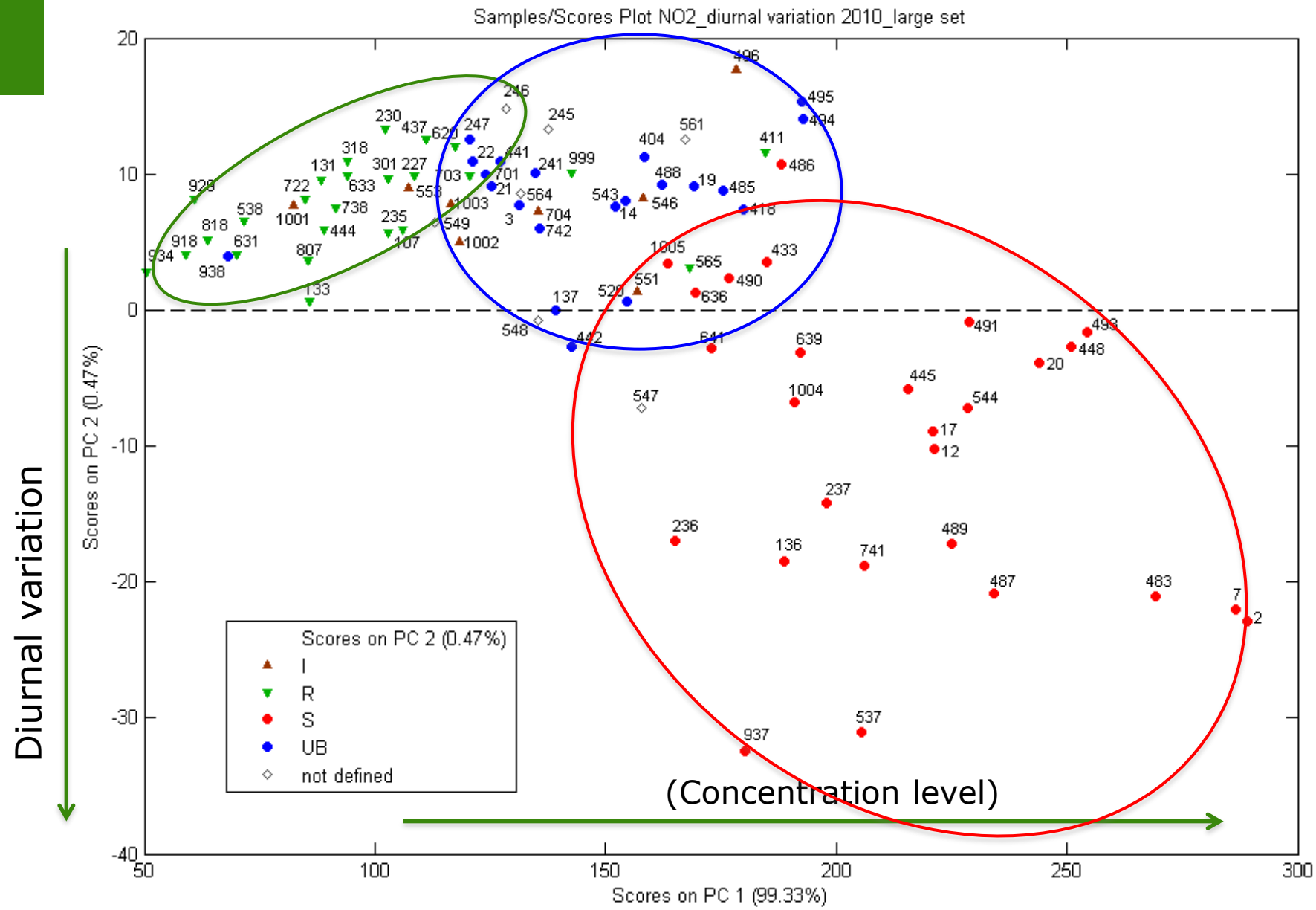
Source: Wikipedia

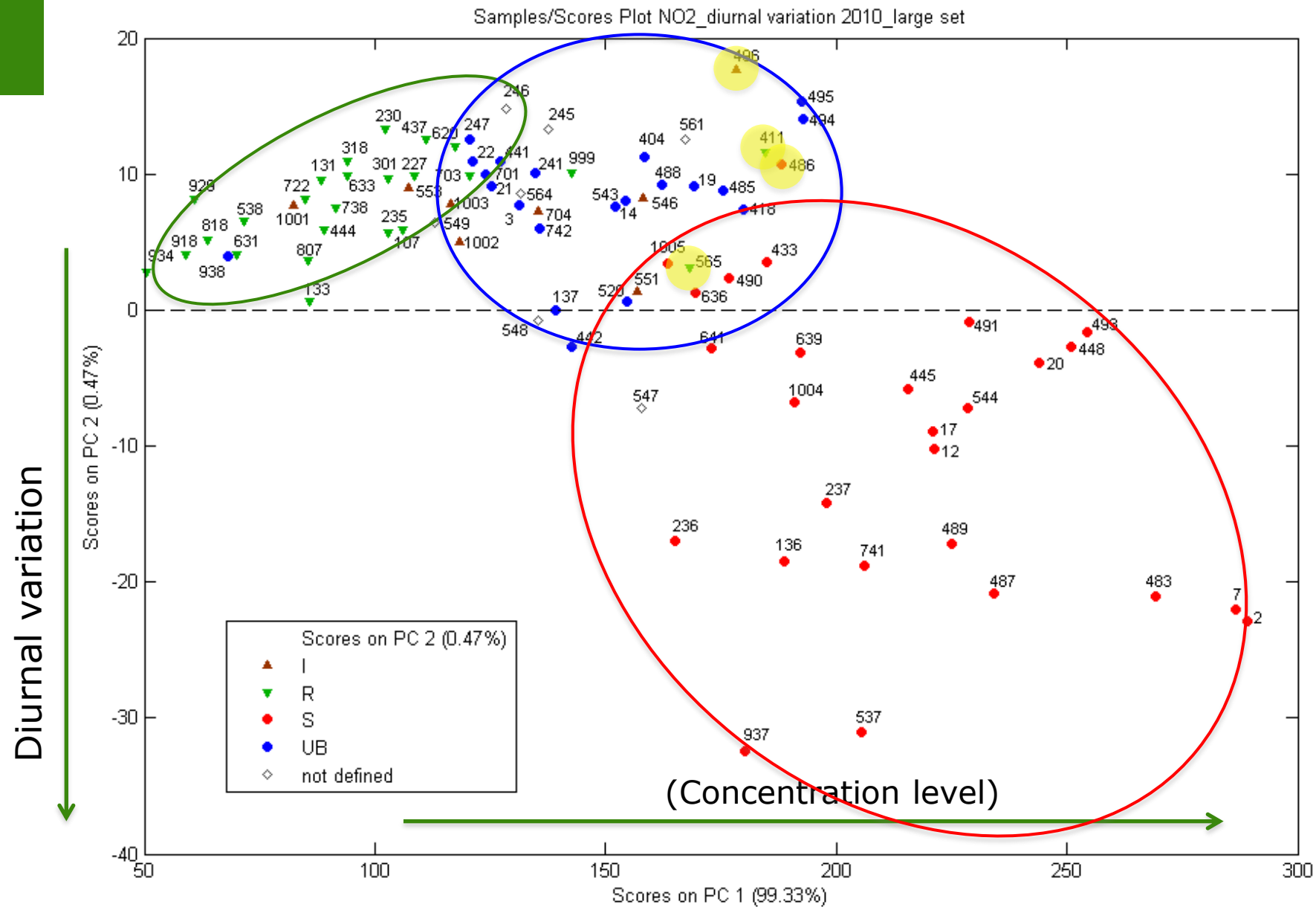
Principal component analysis

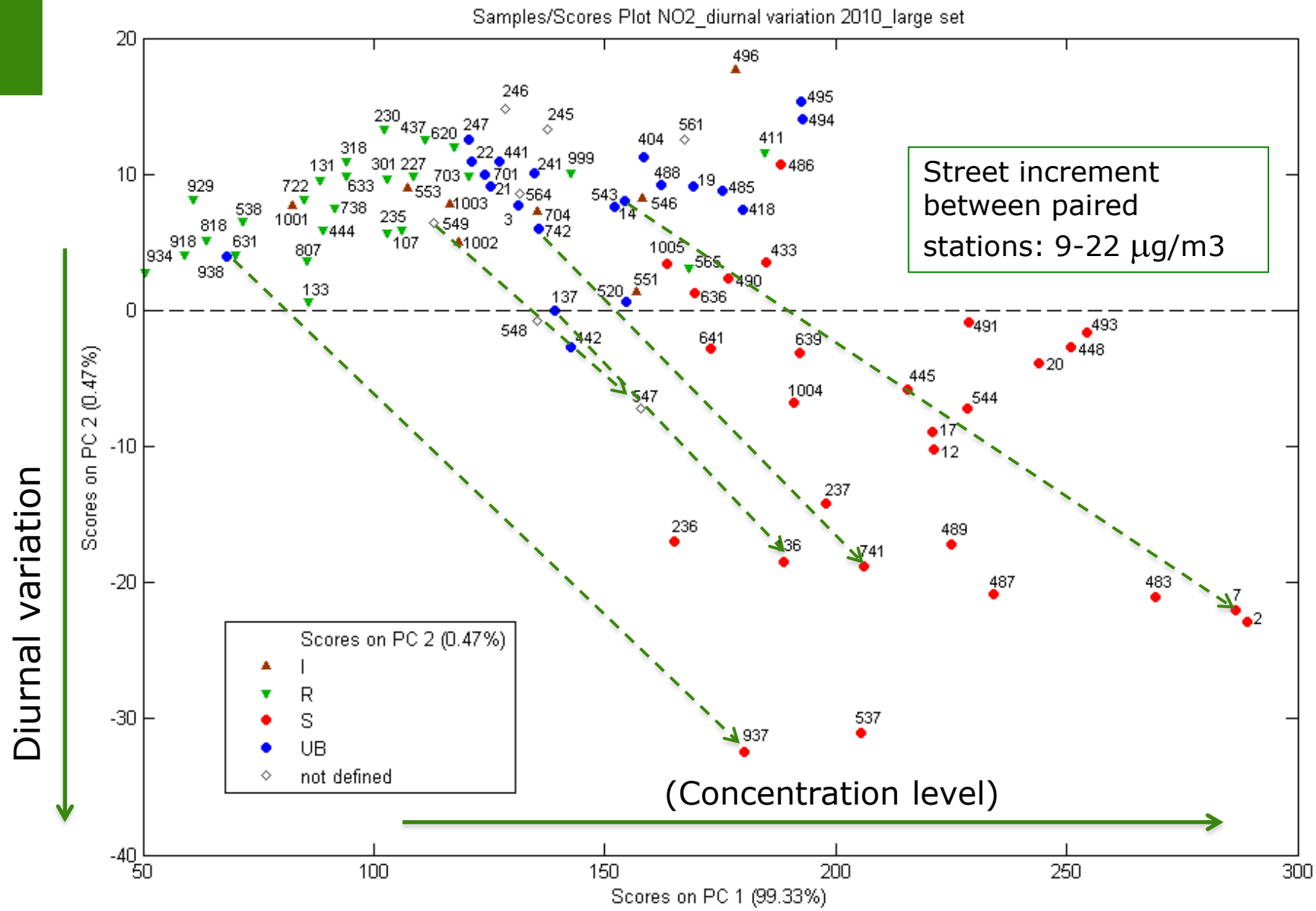


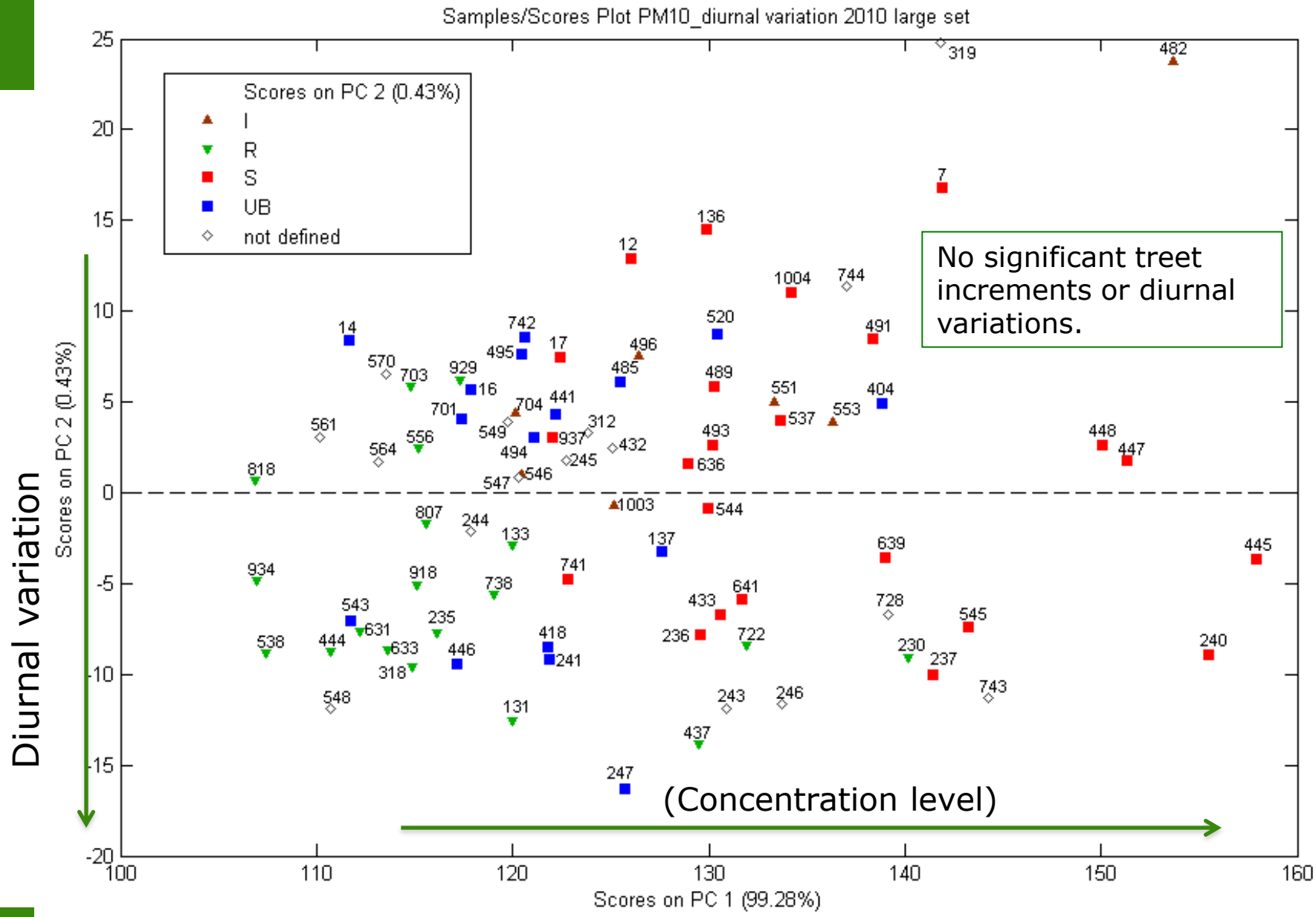
Source: Wikipedia

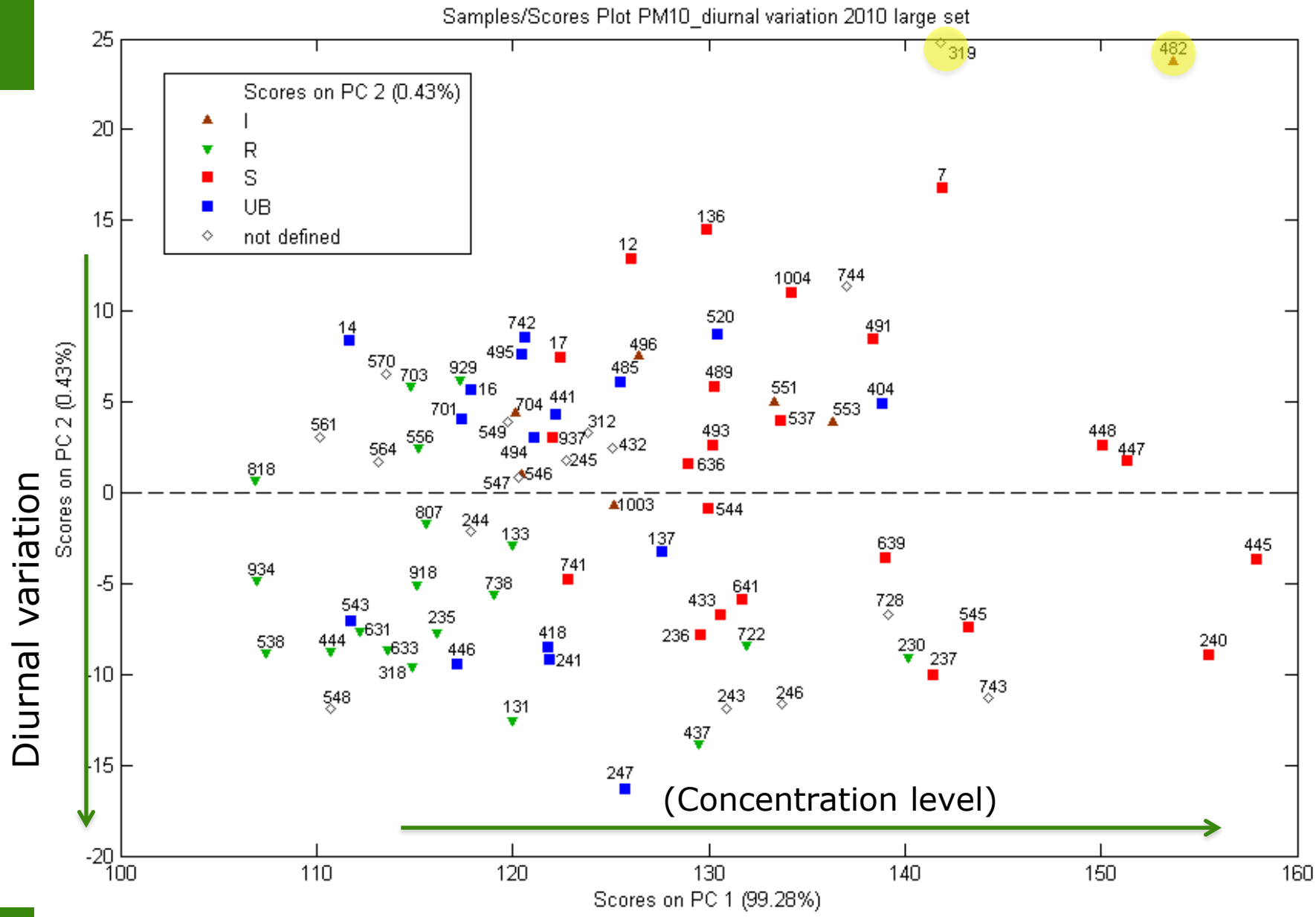

Samples/Scores Plot NO2_diurnal variation 2010_large set
Scores on PC 2 (0.47%)
Scores on PC 1 (99.33%)
Diurnal variation
(Concentration level)
Scores on PC 2 (0.47%)
I
R
S
UB
not defined


Samples/Scores Plot NO2_diurnal variation 2010_large set
Diurnal variation
(Concentration level)
Scores on PC 2 (0.47%)
Scores on PC 1 (99.33%)
Scores on PC 2 (0.47%)
I
R
S
UB
not defined


Samples/Scores Plot NO2_diurnal variation 2010_large set
Diurnal variation
Scores on PC 2 (0.47%)
Scores on PC 1 (99.33%)
(Concentration level)
Scores on PC 2 (0.47%)
I
R
S
UB
not defined
20
10
0
-10
-20
-30
-40
50
100
150
200
250
300


Samples/Scores Plot NO2_diurnal variation 2010_large set
Scores on PC 2 (0.47%)
Scores on PC 1 (99.33%)
Diurnal variation
(Concentration level)
Street increment between paired stations: 9-22 µg/m3
Scores on PC 2 (0.47%)
I
R
S
UB
not defined


Samples/Scores Plot PM10_diurnal variation 2010 large set
Scores on PC 2 (0.43%)
I
R
S
UB
not defined
Diurnal variation
Scores on PC 2 (0.43%)
Scores on PC 1 (99.28%)
(Concentration level)
No significant treet increments or diurnal variations.


Samples/Scores Plot PM10_diurnal variation 2010 large set
Scores on PC 2 (0.43%)
I
R
S
UB
not defined
Diurnal variation
Scores on PC 2 (0.43%)
Scores on PC 1 (99.28%)
(Concentration level)

PCA for NO2 and PM10

- Diurnal variation and concentration for NO2 and PM10.
- Relatively clear groups of similar stations along the lines rural, urban, street.
- Several outliers.
- PM10 data doesn’t show distinct clusters of street and urban background or traffic stations.

Concluding remarks

- Principal Components Analysis seems a valuable (additional) tool to find patterns in groups of monitoring stations and to classify the types.

- The RIVM would like to find out if other institutes also use PCA and what their results are.

- Please email: joost.wesseling@rivm.nl

THANK YOU