Fisica dell'atmosfera e dispersione degli inquinanti

Modelli a recettore

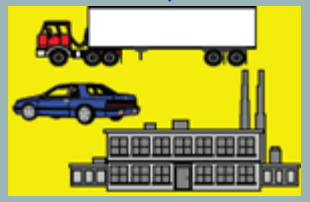
http://www.ge.infn.it/~prati

IL PROBLEMA DEL «SOURCE APPORTIONMENT»
DUE APPROCCI COMPLEMENTARI

Modelli a RECETTORE

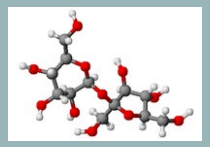
Analisi matematica per individuare le «sorgenti» degli inquinanti

Misura della concentrazione di inquinanti in una zona (il «recettore»)

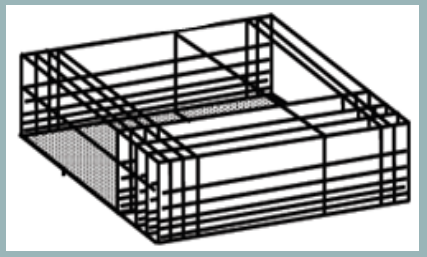


Impatto sulla qualità dell'aria di ciascuna «sorgente»

Modelli chimici di TRASPORTO



Calcolo delle trasformazioni chimiche e del trasposto in atmosfera



Calcolo di mappe di concentrazione degli inquinanti in una rappresentazione schematica del territorio

MODELLI «SOURCE-ORIENTED» E «RECEPTOR-ORIENTED»

Modelli Receptor-oriented

(dal sito di campionamento/osservazione alle sorgenti)

Modelli diagnostici: dalla misura del **PM** e della sua composizione in un sito (il recettore) si identificano le sue «sorgenti» e il loro contributo al PM e/o ai suoi componenti

Modelli Source-oriented

(dalla sorgente al punto di interesse)

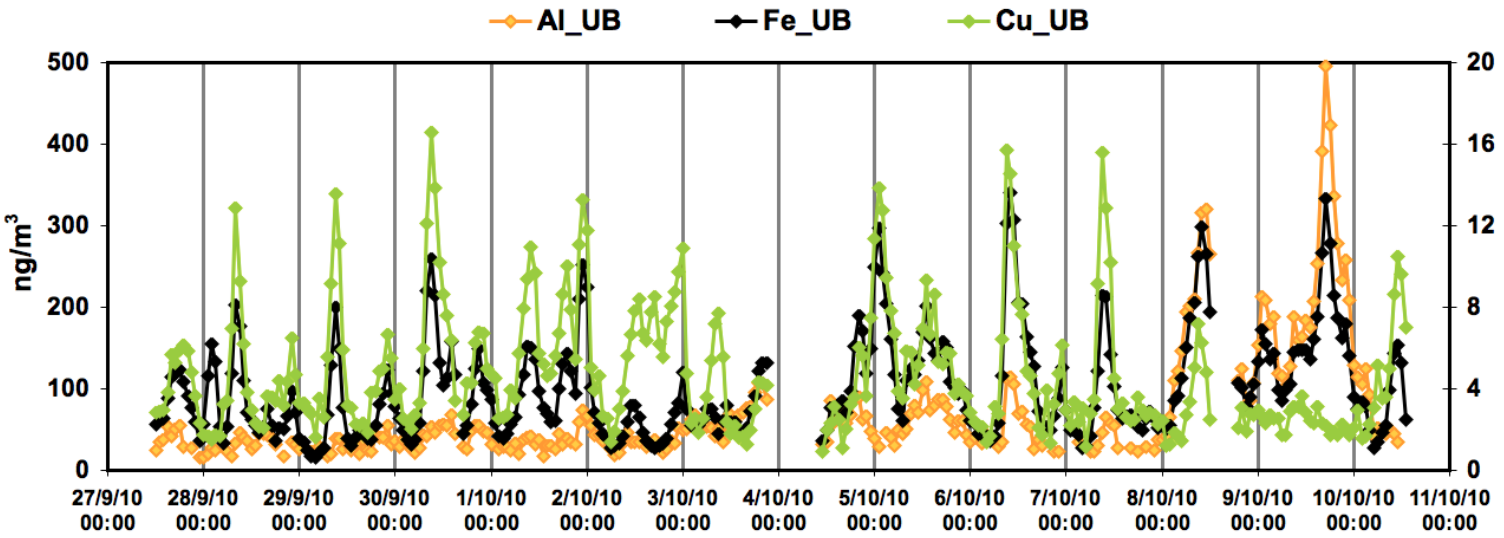
Modelli predittivi: dalla conoscenza delle sorgenti e grazie a modelli di dispersione atmosferica, si calcolano mappe di concentrazione di **PM** (e/o sue componenti) **e gas** sul territorio

LE BASI DEI MODELLI A RECETTORE

I componenti misurati del PM posso essere emessi da una o più sorgenti:

Marker = chemical species which is characteristic of a source

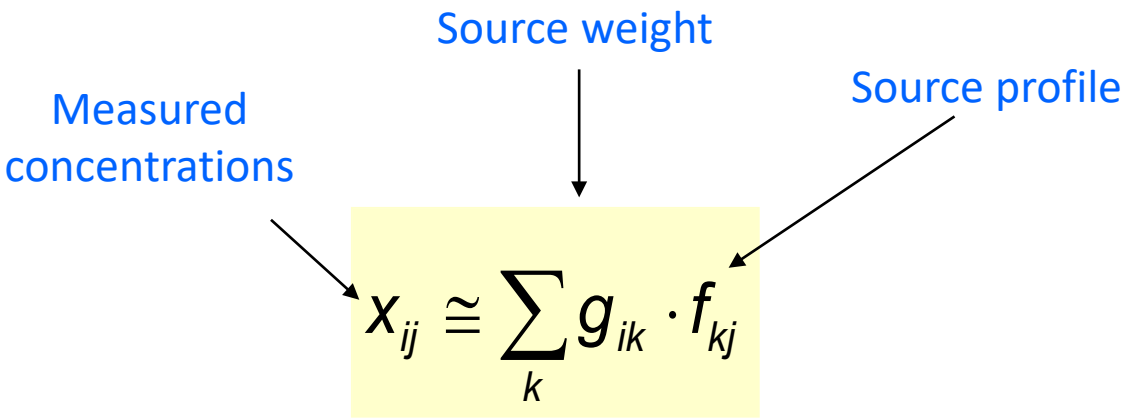
Tracer = unique marker (produced by only one source)



Dal punto di vista matematico, le concentrazioni degli elementi/specie misurati sono variabili eventualmente correlate (i.e. non indipendenti) mentre le sorgenti sono variabili fittizie/sintetiche che originano però la variabilità e le correlazioni delle/tra le singole specie. Si tratta quindi di eseguire una FACTOR ANALYSIS che intende appunto identificare variabili sintetiche che conservino l'informazione espressa dalla variabilità di ciascuna componente del PM.

MASS BALANCE EQUATION

Tutti i modelli a recettore cercano di risolvere l'equazione che esprime la conservazione della massa:



x_{ij} = *concentration of element j in sample i*

g_{ik} = *contribution of source k in sample i*

f_{kj} = *fraction of element j in the PM produced by source k*

Il modello a recettore cerca di ottenere la matrice dei coefficienti $\boldsymbol{g_{ik}}$ (pesi/emissioni di una sorgente) a partire dalle concentrazioni misurate $\boldsymbol{x_{ij}}$. Si tratta quindi di un problema inverso: in alcuni casi si assume di conoscere i profili di emissione delle sorgenti ($\boldsymbol{f_{kj}}$), in altri questi sono ottenuti direttamente dal modello.

IPOTESI ALLA BASE DEI MODELLI A RECETTORE

$$x_{ij} \cong \sum_k g_{ik} \cdot f_{kj}$$

Tutte le sorgenti principali (ovvero che danno un contributo significativo alla concentrazione di PM) vengono individuate

- La composizione chimica del PM deve essere misurata/conosciuta nel modo più complete possibile ed includere markers e tracers delle diverse sorgenti.

I profili emissivi delle sorgenti sono constanti nel tempo (almeno durante il period di campionamento/osservazione)

- Per sorgenti variabili (e.g. cicli industriali che non sono rigidamente fissati) non si può che individuare un valore medio con eventuale impossibilità di riconoscere le reali sorgenti.

[Watson et al., Chemosphere, 2002]

IPOTESI ALLA BASE DEI MODELLI A RECETTORE

$$x_{ij} \cong \sum_k g_{ik} \cdot f_{kj}$$

I profili emissivi delle sorgenti sono constanti nello spazio ovvero non si modificano durante il trasporto del PM dalla sorgente al recettore

- I Marker/Tracers dovrebbero essere chimicamente stabili (e.g.: elementi)→ problemi con traccianti organici
- Notevoli problemi con gli aerosol secondari

I profili emissivi devono essere linearmente indipendenti e le sorgenti non allineate lungo una stessa direzione.

- Sorgenti con profili simili (e.g. dust desertica e di origine locale) possono risultare non distinguibili. In tale case sia la speciazione chimica che la risoluzione temporale del campionamento devono essere affinati.

[Watson et al., Chemosphere, 2002]

IL PROBLEMA DEGLI AEROSOL SECONDARI

- Costituiti principalmente da ioni solfato, nitrato, ammonio e carbonio organico secondario (SOC e VOC), concentrate nelle frazioni PM2.5 e PM1.
- Hanno concentrazioni in atmosfera che dipendono principalmente da quelle dei gas precursori (NO_x, SO_2, ...) e di altre specie gassose reattive (O_3 e radicale OH) e dalle condizioni atmosferiche (umidità, temperatura, irradianza, etc.): legame con le sorgenti di emissioni primarie sostanzialmente non individuabile.
- Largamente impattano sulla proporzionalità tra PM emesso e rilevato in atmosfera.

Conseguentemente il PM secondario viene generalmente apporzionato nelle sue componenti principali (nitrati e solfati secondari, carbonio organico) senza la possibilità di risalire alle sorgenti primarie che hanno emesso i precursori del PM secondario.

Uno dei primi tentativi:

M. C. Bove et al., (2014). An integrated PM2.5 source apportionment study: Positive Matrix Factorisation vs. the Chemical Transport Model CAMx. ATMOSPHERIC ENVIRONMENT, vol. 94, p. 274-286, doi: 10.1016/j.atmosenv.2014.05.039

MODELLI DIAGNOSTICI NON AUTOSUFFICIENTI

Before

Data base "production"

- Well thought choice of: sampling sites, number of samples, time resolution, PM fraction, collecting substrates, analysed chemical species

Data base validation and preliminary analysis

- Study of temporal series, correlations, scatterplots, …
- Comparison among sampling sites
- Comparison with meteorological data (wind roses, etc.)

After

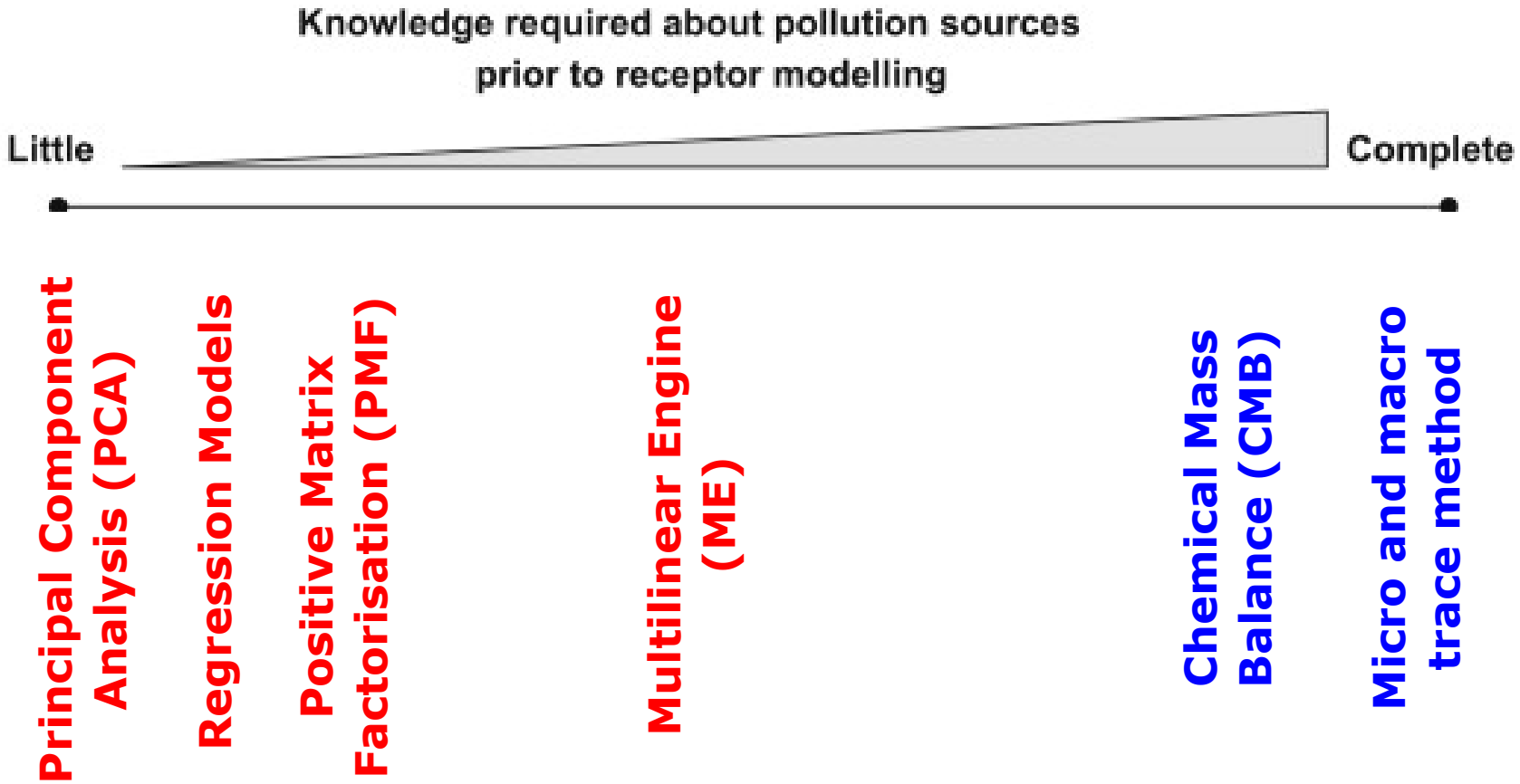
Critical analysis of model results

- Comparison with literature profiles
- Are the source time trends reasonable (also in relation to their location and meteorology) ?

Have a look to: European Guide on Air Pollution Source Apportionment with Receptor Models

[Watson et al., Chemospsphere, 2002]

CLASSIFICAZIONE DEI MODELLI A RECETTORE



ONE-SAMPLE MODELS

$$x_{ij} \cong \sum_k g_{ik} \cdot f_{kj}$$

Independent analysis on each single sample

$$\begin{cases} x_1 \cong g_1 \cdot f_{11} + g_2 \cdot f_{21} + \ldots + g_P \cdot f_{P1} \\ \ldots \\ x_M \cong g_1 \cdot f_{1M} + g_2 \cdot f_{2M} + \ldots + g_P \cdot f_{PM} \end{cases}$$

M species - P sources

(M equations)

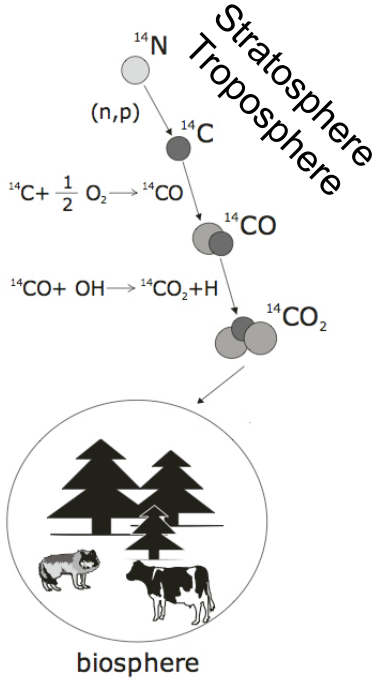
not enough equations for determining both F and G: source profiles (F) are needed

- Trace element method (es. ^{14}C, levoglucosan)

  Example: Biomass burning = f x [levoglucosan]

- Chemical Mass Balance (CMB)

APPORZIONAMENTO DEL PM CARBONIOSO CON ^{14}C

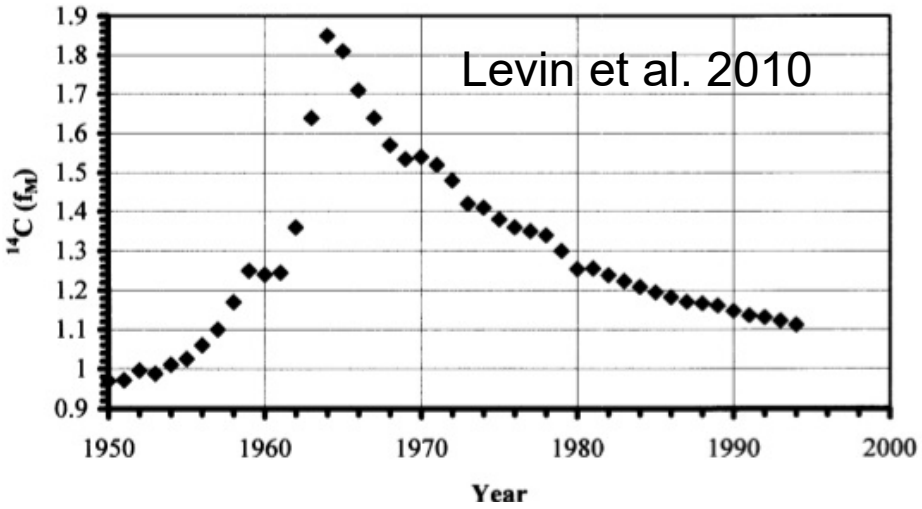


- Rare isotope ($^{14}C/^{12}C \sim 1.2*10^{-12}$), instable ($\tau_{1/2} \sim 5700$ y)
- By convention ^{14}C concentration is defined as:

Fraction of modern carbon

$$f_m = \frac{(^{14}C/^{12}C)_{sample}}{(^{14}C/^{12}C)_{AD1950}}$$

- Fossil material: $f_{m,fossil} = 0$
- Modern material: $f_{m,mod} \gtrsim 1$



Per il PM carbonioso:

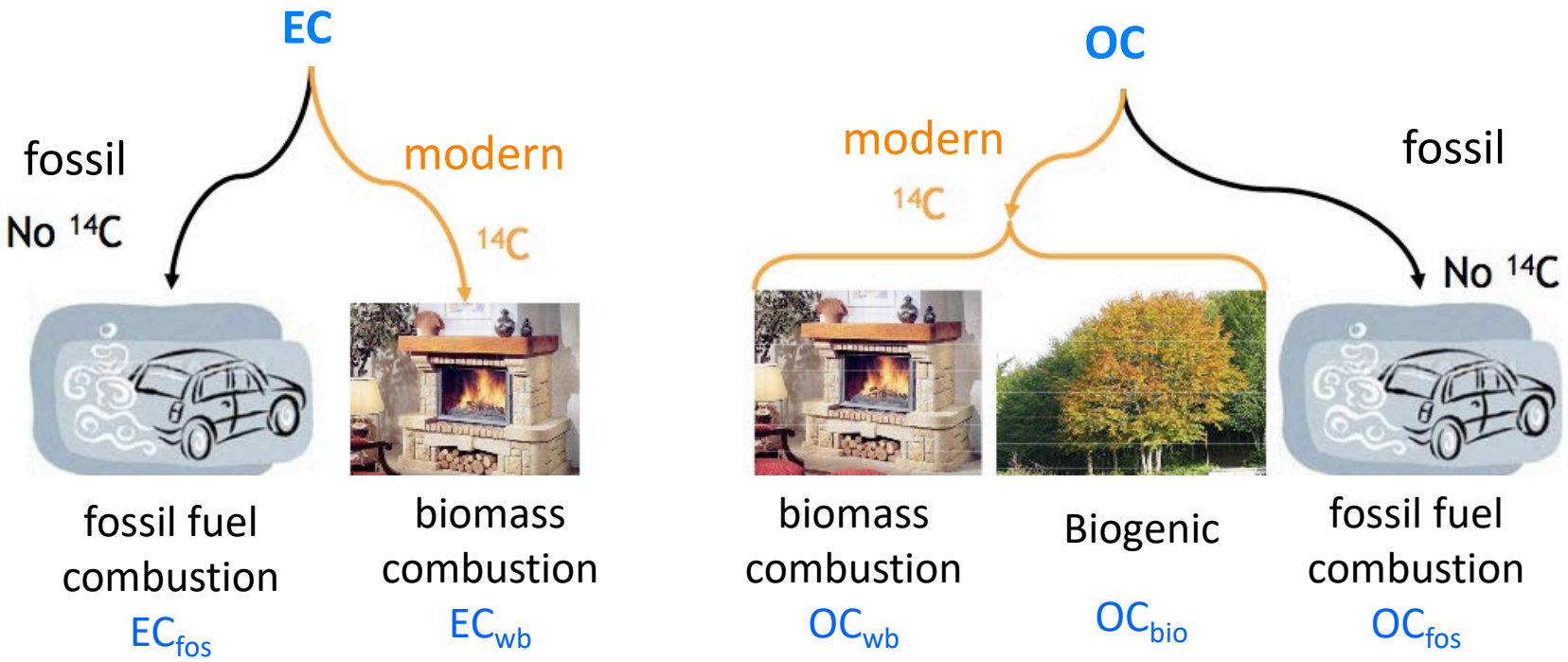
$$TC_{tot} = TC_{fossil} + TC_{mod}$$

$$TC_{tot} \cdot f_m(TC) = TC_{fossil} \cdot f_{m,fossil} + TC_{mod} \cdot f_{m,mod}$$

$$TC_{mod} = TC_{tot} \frac{f_m}{f_{m,mod}}$$

APPORZIONAMENTO DI EC ED OC CON ^{14}C

Per EC sono possibili 2 sorgenti mentre sono 3 quelle per OC:



$(EC/OC)_{wb}$

APPORZIONAMENTO DI EC ed OC CON ^{14}C

$$EC_{wb} = \frac{EC \cdot f_m(EC)}{f_{m,wb}}$$

$$EC_{fossil} = EC - EC_{wb}$$

$f_{m,fossil} = 0$

$f_{m,wb} = 1.083$ (Minguillon et al., 2011)

$(OC/EC)_{wb} = 5.5 \pm 1.2$

By PMF (Bernardoni et al., 2011)

$$\boldsymbol{OC_{wb} = \frac{EC_{wb}}{\left(EC/OC\right)_{wb}}}$$

$$\boldsymbol{OC_{bio} = OC\frac{f_m(OC)}{f_{m,mod}} - OC_{wb}}$$

$$\boldsymbol{OC_{fos} = OC - OC_{wb} - OC_{bio}}$$

$f_{m,bio} = 1.040 \pm 0.004$
(Levin et al., 2008)

$f_{m,mod}$ = average between $f_{m,bio}$ and $f_{m,wb}$

[Bernardoni et al. JAS 2013]

ONE-SAMPLE SOURCE APPORTIONMENT

(example from Seinfield and Pandis, 2006. Atmospheric Chemistry and Physics)

$$x_{ij} \cong \sum_k g_{ik} \cdot f_{kj}$$

Total concentration of particulate Fe at a site can be considered to be the sum of contributions from a number of independent sources:

$$Fe_{tot} = Fe_{soil} + Fe_{traffic} + Fe_{coal} + ...$$

Receptor site: a rather simple scenario at a rural site is assumed:

Measured concentration values at the receptor site:

$$PM10 = 32\ \mu g\ m^{-3};\ Fe = 3.84\ \mu g\ m^{-3}\ ;\ Si = 2.58\ \mu g\ m^{-3}$$

Hp: two major sources contribute to PM concentration at the site with the following, simple, "emission" profiles:

coal-fired power plant, COAL: Si= 10 mg/gPM10; Fe= 150 mg/g PM10

soil resuspension, SOIL: Si= 200 mg/g PM10; Fe=32mg/g PM10

ONE-SAMPLE SOURCE APPORTIONMENT

$$x_{ij} \cong \sum_k g_{ik} \cdot f_{kj}$$

Neglecting Si and Fe contributions from other sources:

$$Si_{tot} = Si_{SOIL} + Si_{COAL}$$

$$Fe_{tot} = FE_{SOIL} + Fe_{COAL}$$

and:

$$PM10 = PM10_{SOIL} + PM10_{COAL} + PM10_{OTHER}$$

If the composition of particles does not change during their transport from the sources to the receptor, we use the initial composition of the emissions, and we obtain:

$$Si_{tot} = PM10_{SOIL} * 0.2 + PM10_{COAL} * 0.01$$

$$Fe_{tot} = PM10_{SOIL} * 0.032 + PM10_{COAL} * 0.15$$

Solving the system of 2 equations with 2 unknown variables (PM_{SOIL} and PM_{COAL}), we obtain the solution:

$$PM10_{COAL} = 18\ \mu g\ m^{-3}\ (\ 56.2\%\ PM10)$$

$$PM10_{SOIL} = 12\ \mu g\ m^{-3}\ (37.5\%\ PM10)$$

$$PM10_{OTHER} = 2\ \mu g\ m^{-3}\ (6.3\%\ PM10)$$

ONE-SAMPLE SA: CHEMICAL MASS BALANCE (CMB)

$$x_{ij} \cong \sum_k g_{ik} \cdot f_{kj}$$

CMB combines the chemical composition of PM measured at the sources (f_{kj}) and at the receptor (X_j) to quantify the source contributions (g_k)

CMB is then a regression problem.

However, there are errors in both variables (f_{ki} and X_j), so it is necessary to solve the problem using error models

$$X_j = \sum_{k=1}^{p} g_k f_{kj} + e_j$$

Mass balance equation written for just one sample

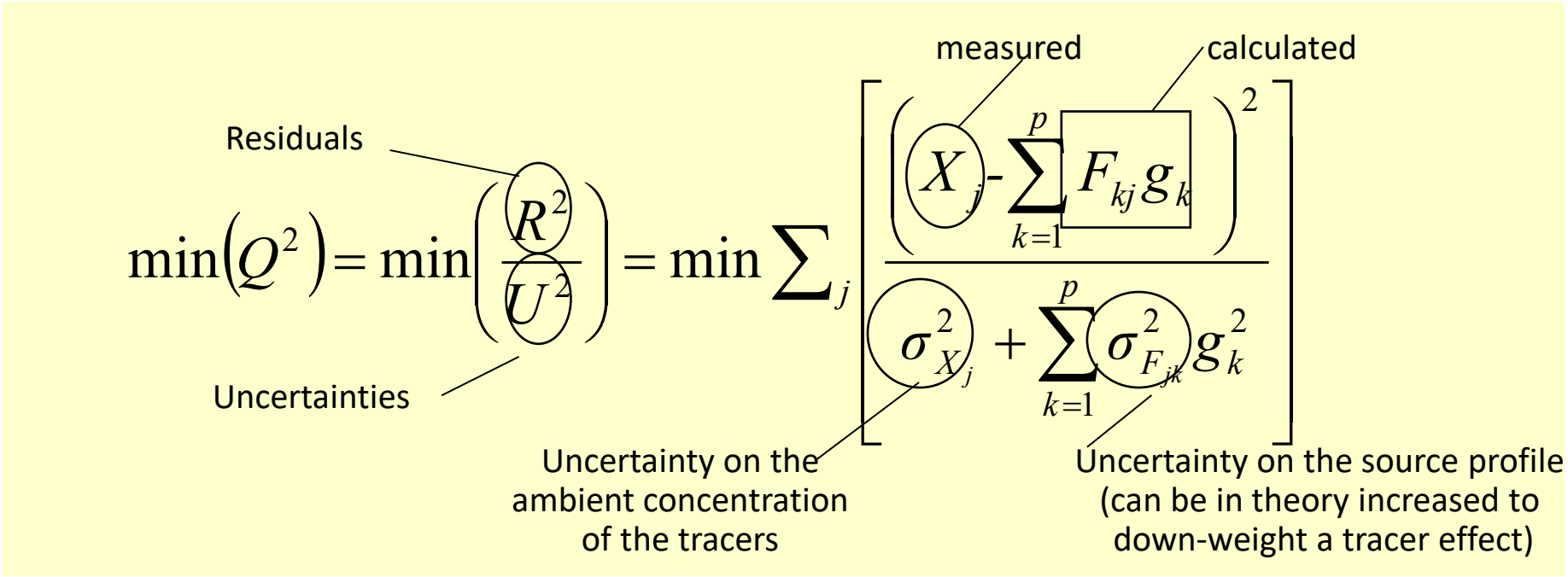
SOLUZIONE DELLA CMB

Effective variance weighted least-squares solution

Soluzione efficace con i minimi quadrati ponderati per la varianza

(Watson et al., AE 1984), adopted by EPA, and incorporated in the CMB8.2 software

The solution minimize differences between calculated and measured values



One solution, which contains the effects of random uncertainties in both receptor concentrations and source compositions

CMB- REQUIREMENTS/ASSUMPTIONS

Selection of tracers and source profiles (as input data)

Tracers:

Stable during transport *(low volatile and non-reactive)*

Accurately determined at the receptor site

Reported for all source profiles considered in the model

Source profiles:

Average well-known representative source profiles *(pertinent to the study location and time)*

All major primary sources of tracers are included *(no missing sources)*

The number of sources is less or equal to the number of tracers (**m sources < j chemical species**)

There is no relationship among the source uncertainties **(σ_{Fkj})**

Measurements uncertainties **(σ_{xj})** are random, uncorrelated and normally distributed

CMB WEAKNESSES/LIMITS

Sources selected in advance → *CMB cannot identify new unknown sources*

Source profiles: are they really representative? → *up-to date and locally relevant profiles required for CMB modeling may not be available*

Source profiles: enough measurement data are available? → *profiles often have not been measured on representative sources, or on enough different sources: the propagated standard errors associated with CMB source-contribution estimates underestimate the true uncertainties of the apportionment in these cases.*

Differences between profiles for the same source category → *Many profiles may be available for main sources (e.g. traffic, biomass burning) with great differences.* → ***composite profiles and sensitivity tests***

COMPOSITE PROFILES and SENSITIVITY TESTS

Source profiles are intended to be representative of a category of source (e.g traffic) rather than individual emitters (e.g. individual vehicles, whit emission changing according to many factors: fuel alimentation, driving conditions, speed, age……)

Composite profiles: combination of individual profiles representing the same source category measured by sampling emission from a variety of single emitters or small groups of emitters (each individual profile is formed from individual samples: A ± SD)

The simplest composite profile consists of the: **AVERAGE ± SD** of abundances for all individual profiles within a group

Sensitivity tests: Test different profiles and combination of profiles → Assess the sensitivity of the CMB results to the selection of different profiles (quality control check criteria: R^2, χ^2 , etc.)

IN A SENSITIVITY STUDY:

The CMB model is run more times: each time a different profile for a certain source category (for which the sensitivity test is being performed) is used as input data in combination with the other fixed sources

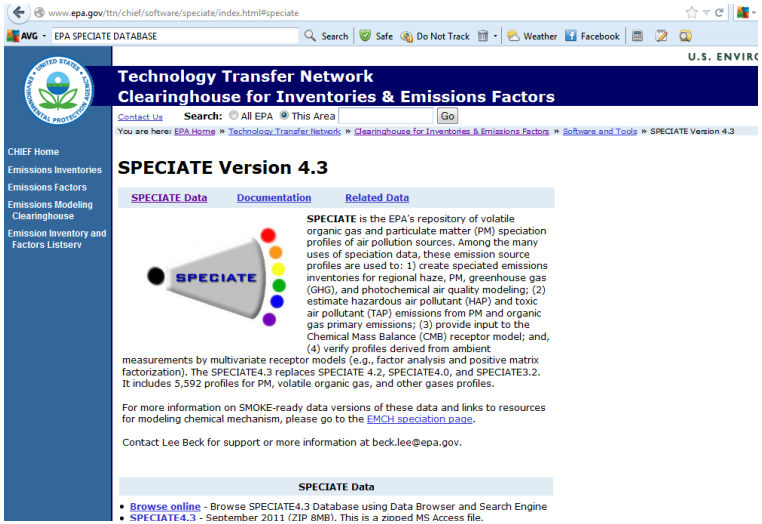
A range of source contribution result is derived from all the tested possible solutions

The average (and/or alternatively the best estimate) and standard deviation of all CMB runs that met the goodness-of fit criteria can be reported as the estimated source contribution and related standard deviation ($g_k + \sigma_{gk}$)

AVAILABILITY OF SOURCE PROFILES

EPA SPECIATE database (US)

http://www.epa.gov/ttnchie1/software/speciate/



VOCs and PM speciation profiles of air pollution sources

SPECIEUROPE database (EU)

http://source-apportionment.jrc.ec. europa.eu/Specieurope

MULTIVARIATE MODELS

$$x_{ij} \cong \sum_k g_{ik} \cdot f_{kj}$$

Analysis on multiple samples: the source apportionment is simultaneously extracted from the whole data matrix. The increased complexity is rewarded by the quantity of obtained information (not only the weights but also the source profiles)

The wide family of all "Factor Analysis" techniques: Target Transformation Factor Analysis (TTFA), **Principal Component Factor Analysis (PCFA), Positive Matrix Factorisation (PMF)**, Multilinear Engine (ME)

How many samples?

As a general rule [Henry et al. 84]: N > 30 + (M+3)/2

FACTOR ANALYSIS

$$x_{ij} \cong \sum_k g_{ik} \cdot f_{kj}$$

Data matrix (measured concentrations) N x M

Factor Scores (source contributions) N x P

Factor Loadings (source profile) P x M

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & \cdots & x_{1M} \\ x_{21} & x_{22} & \cdots & \cdots & x_{2M} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{N1} & x_{N2} & \cdots & \cdots & x_{NM} \end{pmatrix} \cong \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1P} \\ g_{21} & g_{22} & \cdots & g_{2P} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ g_{N1} & g_{N2} & \cdots & g_{NP} \end{pmatrix} \cdot \begin{pmatrix} f_{11} & f_{12} & \cdots & \cdots & f_{1M} \\ f_{21} & f_{22} & \cdots & \cdots & f_{2M} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ f_{P1} & f_{P2} & \cdots & \cdots & f_{PM} \end{pmatrix}$$

Source chemical profile

Chem. species time trend

Source weigth time trend

con P < M

The original matrix is approximated by the product of 2 smaller matrices

The aim is reducing the data set dimension without reducing too much the information that it contains

RANK P APPROXIMATION

The data matrix (rank M) is approximated by the sum of P (<M) matrices, each of those is of rank one (and represents a single source):

$$X \cong X^1 + X^2 + \ldots + X^P$$

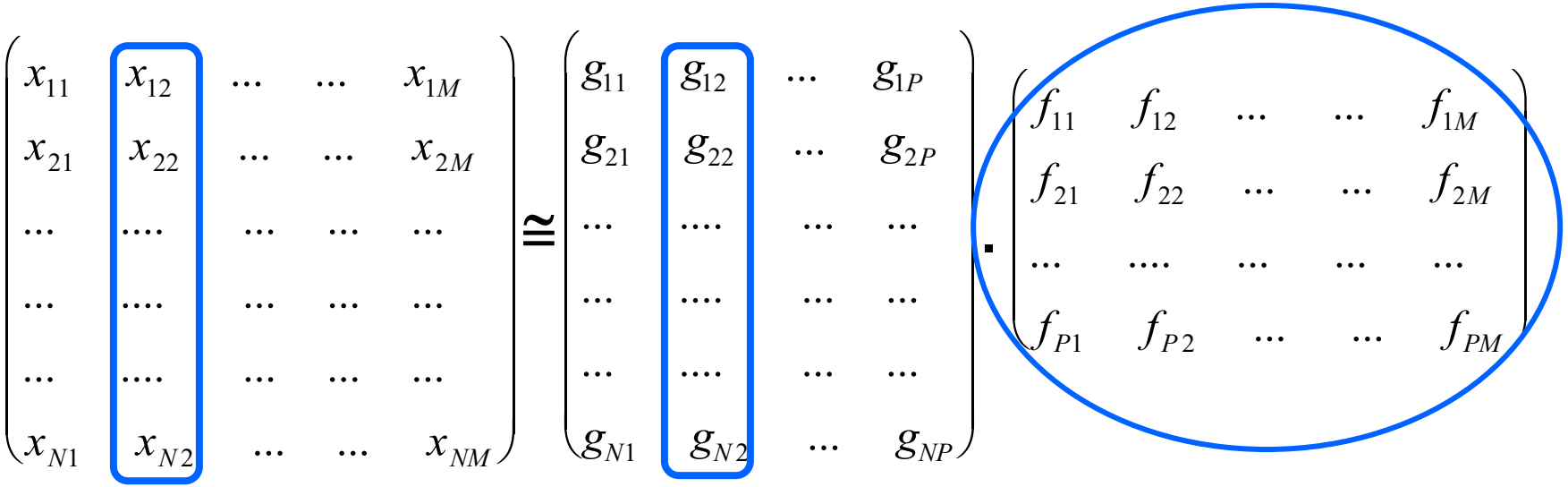
rank= number of linear independent columns

For example, X^2 (due to source number 2) is:

$$\begin{pmatrix} x^2_{11} & x^2_{12} & \ldots & x^2_{1M} \\ x^2_{21} & x^2_{22} & \ldots & x^2_{2M} \\ \ldots & \ldots. & \ldots & \ldots \\ \ldots & \ldots. & \ldots & \ldots \\ x^2_{N1} & x^2_{N2} & \ldots & x^2_{NM} \end{pmatrix} \cong \begin{pmatrix} 0 & g_{12} & \ldots & 0 \\ 0 & g_{22} & \ldots & 0 \\ \ldots & \ldots. & \ldots & \ldots \\ \ldots & \ldots. & \ldots & \ldots \\ 0 & g_{N2} & \ldots & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 & \ldots & \ldots & 0 \\ f_{21} & f_{22} & \ldots & \ldots & f_{2M} \\ \ldots & \ldots. & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \ldots & 0 \end{pmatrix}$$

These “source 2” G and F have only one column/line ≠ 0, and are obviously of rank 1, but also X^2 is of rank 1 as all its columns are linearly proportional to the second column of G.

CHANGE OF VARIABLES



M original variables: measured concentrations of the chemical species

$P < M$ new variables
(linear combinations of the old ones):
"components" or "factors"

The new variables explain most of the data variability (i.e. of most the total variance of the dataset) and/or reconstruct most the measured concentrations.

In some cases, they are constrained to be uncorrelated (PCA).

FA MODELS

Principal Component (Factor) Analysis

based on orthogonal principal axes decomposition.

It has least square properties, but is not weighted with uncertainties and variables are usually standardised (loss of information on the origin of the scale of variables, use of non-negativity constrains is not possible)

Positive Matrix Factorisation (PMF)

least squares fit weighted with uncertainties, with non-negativity constrains

Multilinear Engine (ME)

weighted least squares fit with non-negativity and other user-defined constrains

PRINCIPAL COMPONENT ANALYSIS (PCA)
Geometric approach

Experimental data are points in the R^m space of the M measured chemical species (called "space of the objects").

PCA search for the best-fitting set of orthogonal new axes

(i.e axes which the cloud of points are closest to).

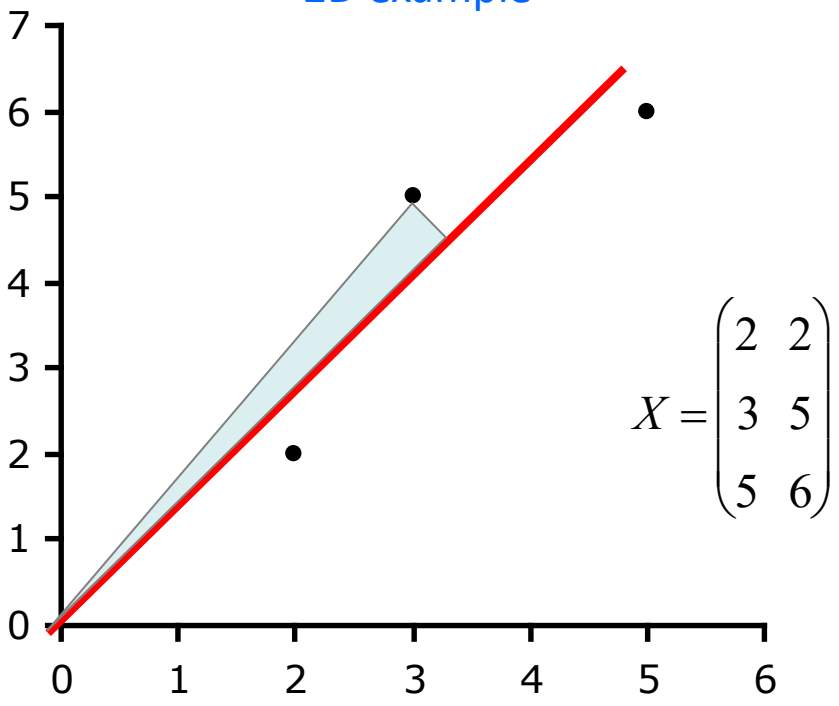
The first axis is obtained minimizing the sum of the square distances between points and axis,

that is equivalent to maximizing the sum of the square projections onto the axis (Pitagora),

i.e. maximizing the variance (or spread) of the point when projected on this axis.

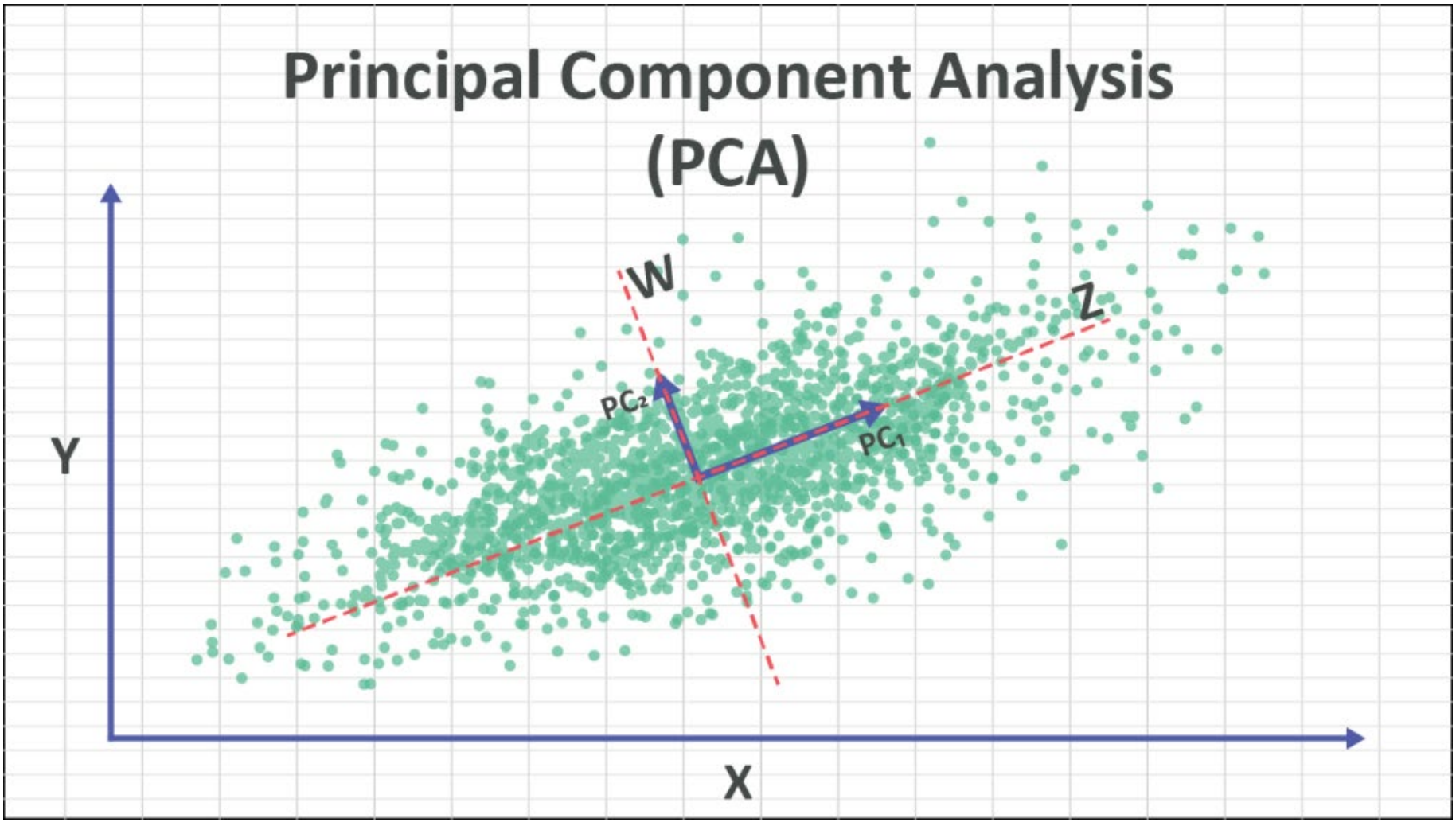
The second axis is obtained by the same approach but in the subspace which is orthogonal to the first axis, and so on.

2D example



$$X = \begin{pmatrix} 2 & 2 \\ 3 & 5 \\ 5 & 6 \end{pmatrix}$$

AN EXAMPLE WITH REALISTIC DATA



Vedi approfondimenti nel materiale didattico nella pagina del corso

CALCULATION OF THE FIRST COMPONENT

$$X\cdot\vec{u}=\begin{pmatrix} x_{11} & \dots & x_{1M} \\ \dots. & \dots & \dots \\ x_{N1} & \dots & x_{NM} \end{pmatrix}\cdot\begin{pmatrix} u_1 \\ \dots \\ u_M \end{pmatrix}$$

projection of X on axis u

We maximize:

$$\| X\cdot\vec{u}\|^2=(X\cdot\vec{u})'(X\cdot\vec{u})=\vec{u}'\cdot X'X\cdot\vec{u}$$

(sum of squares of projections)

with: $\vec{u}'\cdot\vec{u}=1$

Lagrange multipliers:

$$\delta(\vec{u}'\cdot X'X\cdot\vec{u}-\lambda(\vec{u}'\cdot\vec{u}-1))=2X'X\cdot\vec{u}-2\lambda\vec{u}=0 \Rightarrow \boxed{X'X\cdot\vec{u}=\lambda\vec{u}}$$

the first axis is an eigenvector of *X'X*

Multiplying by *u'* we find that λ is equal to the quantity to be maximized ($\vec{u}'\cdot X'X\cdot\vec{u}=\lambda$)

$$\boxed{X'X\cdot\vec{u}_1=\lambda_1\vec{u}_1}$$

the first axis is the eigenvector u_1 of *X'X* corresponding to the largest eigenvalue λ_1

$\lambda_1=\vec{u}_1'\cdot X'X\cdot\vec{u}_1$ **represents the sum of squared projections on the first axis and indicates the amount of variance explained by the first axis**

[Murtagh and Heck, Multivariate Data Analysis, Reidel Publishing Company, 1992]

CALCULATION OF THE OTHER COMPONENTS

The second axis is orthogonal to the first: $\vec{u}' \cdot \vec{u}_1 = 0$

$$\Rightarrow \quad \delta(\vec{u}' \cdot X'X \cdot \vec{u} - \lambda_2(\vec{u}' \cdot \vec{u} - 1) - \mu_2(\vec{u}' \cdot \vec{u}_1)) = 2X'X \cdot \vec{u} - 2\lambda_2\vec{u} - \mu_2\vec{u}_1 = 0$$

$$X'X \cdot \vec{u}_2 = \lambda_2\vec{u}_2$$

the second axis is the eigenvector $\underline{u_2}$ of *X'X* corresponding to the second largest eigenvalue λ_2

The eigenvectors of *X'X*, arranged in decreasing order of corresponding eigenvalues, are the principal components: they give the line of best fit to the clouds of points, the plane of best fit, the 3-dimensional hyperplane of best fit, and so on for higher dimensional subspaces of best fit.

Eigenvectors of *XX'* associated with the *P* largest eigenvalues yield the best-fitting (unweighted!) P-dimensional subspace. The sum of the first P eigenvalues is the sum of squared projections of all the points on this subspace and represents the variance explained by the model.

LIMITS OF PCA

Uncertainties on measured data are not taken into account: concentrations of all chemical species are equally weighted independently from the accuracy of their measurements

Anti-correlations/negative source apportionments: some factor loadings may be negative (it means anti-correlation among factors and measured species): to avoid negative apportionments they should be set equal to zero

Collinear sources (in time trends):as components are by definition un-correlated, the model cannot describe real collinear sources (due for example to meteorological factors)

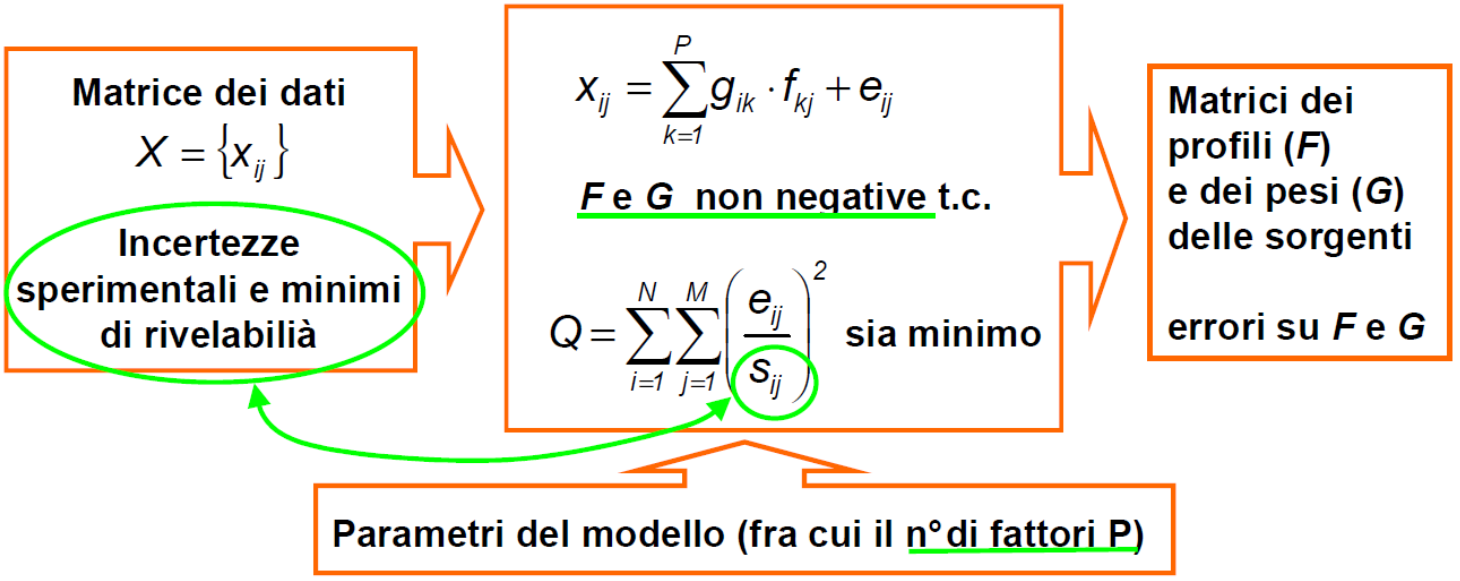
Experimental uncertainties of the model results: the model outputs are given without uncertainty or with uncertainties that do not take into account the experimental uncertainties on input data.

POSITIVE MATRIX FACTORIZATION

Principio di base

- **si parte direttamente dall'equazione del bilancio di massa**
- **tramite un algoritmo iterativo si determina la fattorizzazione con matrici G ed F positive tali da minimizzare lo scarto fra le concentrazioni misurate e quelle ricostruite dal modello**

METODO DEI MINIMI QUADRATI PESATI, CON VINCOLO DI NON NEGATIVITA'



[Paatero and Tapper, Environmetrics, 1994]

POSITIVE MATRIX FACTORIZATION

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1M} \\ x_{21} & x_{22} & \dots & x_{2M} \\ \dots & \dots & \dots & \dots \\ x_{N1} & x_{N2} & \dots & x_{NM} \end{pmatrix}_{NxM}$$

campione

specie chimica

Riepilogo delle matrici

$$x_{ij} = \sum_{k=1}^{P} g_{ik} \cdot f_{kj} + e_{ij}$$

Factor Scores (matrice dei contributi temporali):

$$G = \begin{pmatrix} g_{11} & g_{12} & \dots & g_{1P} \\ g_{21} & g_{22} & \dots & g_{2P} \\ \dots & \dots & \dots & \dots \\ g_{N1} & g_{N2} & \dots & g_{NP} \end{pmatrix}_{NxP}$$

campione

andamento temporale della sorgente

Factor Loadings (matrice dei profili):

$$F = \begin{pmatrix} f_{11} & f_{12} & \dots & f_{1M} \\ f_{21} & f_{22} & \dots & f_{2M} \\ \dots & \dots & \dots & \dots \\ f_{P1} & f_{P2} & \dots & f_{PM} \end{pmatrix}_{PxM}$$

profilo chimico della sorgente

specie chimica

VINCOLI su G: valori positivi

VINCOLI su F: valori positivi e minori di 1

POSITIVE MATRIX FACTORIZATION

Assunzioni per l'apporzionamento

- ***G* non negativa**
  perchè un elemento negativo rappresenterebbe un pozzo di massa anziché una sorgente
- **F non negativa**
  per definizione di profilo

Condizioni direttamente implementate dal modello

- **Limite sui valori di F**
  per definizione di profilo gli elementi di F non possono essere maggiori di 1
- **Limite sui valori di G**
  per ogni campione la somma dei contributi di tutte le sorgenti non può essere maggiore della concentrazione del PM

Condizioni da verificare a posteriori (validazione del modello)

POSITIVE MATRIX FACTORIZATION

Metodo dei minimi quadrati pesati

$$Q = \sum_{i=1}^{N}\sum_{j=1}^{M}\left(\frac{e_{ij}}{s_{ij}}\right)^2$$

Funzione da minimizzare

con $1/s_{ij}$ peso del dato x_{ij}, s_{ij} legato all'incertezza sperimentale del dato stesso

- La possibilità di inserire ciascun dato con il suo peso permette di sfuttare correttamente e efficacemente l'informazione contenuta nel data-set, consentendo anche l'inserimento di dati problematici (casi mancati, casi minori o prossimi al minimo di rivelazione)
- Il peso rende inoltre Q adimensionale ⇒ invarianza per cambiamenti di scala (SCALING OTTIMALE)

Se la stima di s_{ij} è corretta, Q dovrebbe essere governata approssimativamente da una distribuzione chi-quadro (χ^2):

Il valore di aspettazione di Q è dato dal numero di gradi di libertà

POSITIVE MATRIX FACTORIZATION

Dati di input: concentrazioni, incertezze e trattamento dati mancanti e < MDL

$$\text{Rapporto S/N} = \frac{\text{(somma di tutte le concentrazioni > MDL)}}{\text{<MDL> x n°di casi < MDL}}$$

S/N =
- <0.2 variabile cattiva
- 0.2-2 variabile debole
- >2 variabile normale

Criterio di Polissar per calcolare le matrici X ed S:

$$\text{dati noti}: \quad x_{ij} = c_{ij}, \quad s_{ij} = \sigma_{ij} + \frac{MDL_{ij}}{3}$$

$$\text{dati} < \text{MDL}: \quad x_{ij} = \frac{MDL_{ij}}{2}, \quad s_{ij} = \frac{\overline{MDL}_j}{2} + \frac{MDL_{ij}}{3}$$

$$\text{dati mancanti}: \quad x_{ij} = c_j, \quad s_{ij} = 4\tilde{c}_j \quad \text{con } \tilde{c}_j \text{ media geometrica}$$

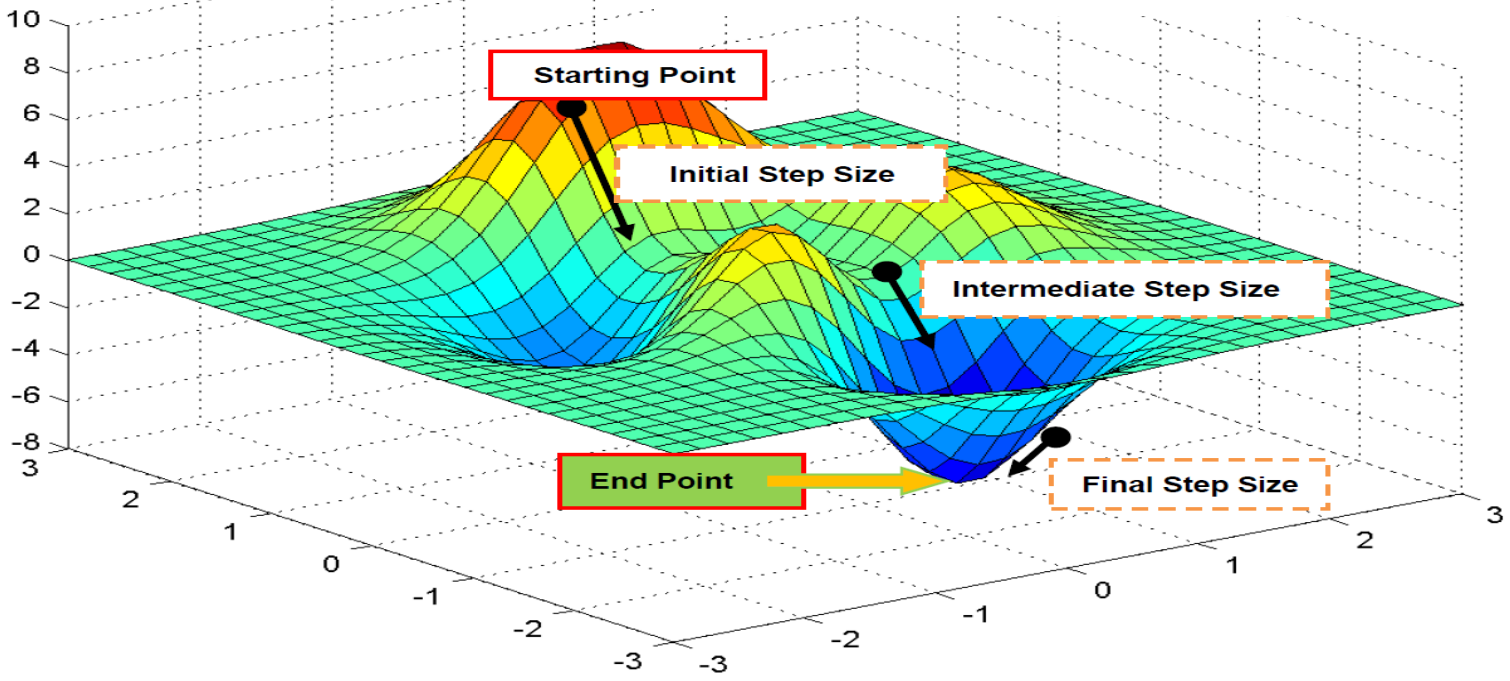
(c_{ij} concentrazione misurata e σ_{ij} incertezza sperimentale)

[Polissar et al., JGR, 1998]

POSITIVE MATRIX FACTORIZATION

Multiple-solution problem

The search for the solution goes from coarser to a finer scale over three levels of iterations.



POSITIVE MATRIX FACTORIZATION

Ambiguità rotazionale

Uno stesso problema di fattorizzazione ammette infinite soluzioni:

$$X = G\,F = G\,T^{-1}\,T\,F = (G\,T^{-1})(T\,F)$$

con: T matrice invertibile (rotazione)

Nella PMF la condizione di non-negatività di *G* ed *F* riduce il numero di soluzioni possibili ma non elimina del tutto l'ambiguità rotazionale.

è necessario selezionare la rotazione ottimale, ovvero la soluzione (*G*, *F*) ottimale, in base ad altri criteri:

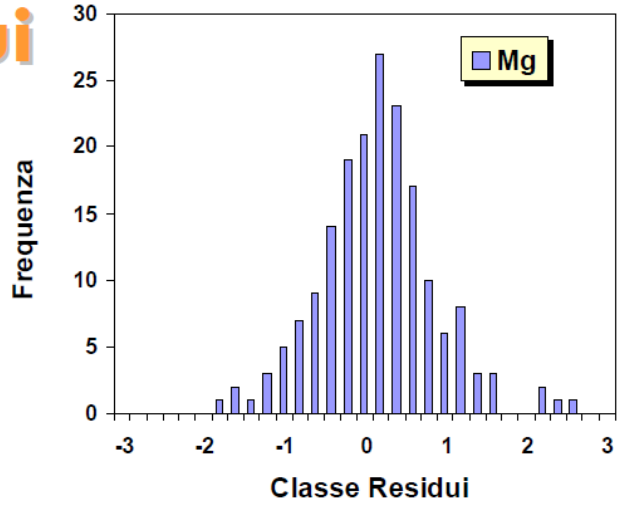
- **Vincoli su *F*: valori minori di 1, ragionevolezza dei profili rispetto a quanto noto (assenza/presenza di specie chimiche in taluni profili, rapporti elementali coerenti con quanto noto, ...)**
- **Vincoli su *G*: somma dei contributi delle sorgenti non superiore alla concentrazione del PM, ragionevolezza degli andamenti e dei contributi delle sorgenti (assenza/presenza di una sorgente in un determinato periodo, ...)**

Valutazione a posteriori esplorando le possibili soluzioni (vedi parametro *FPEAK* in PMF2) e/o imposizione di vincoli a priori (vedi matrici F_{key} e G_{key} in PMF2)

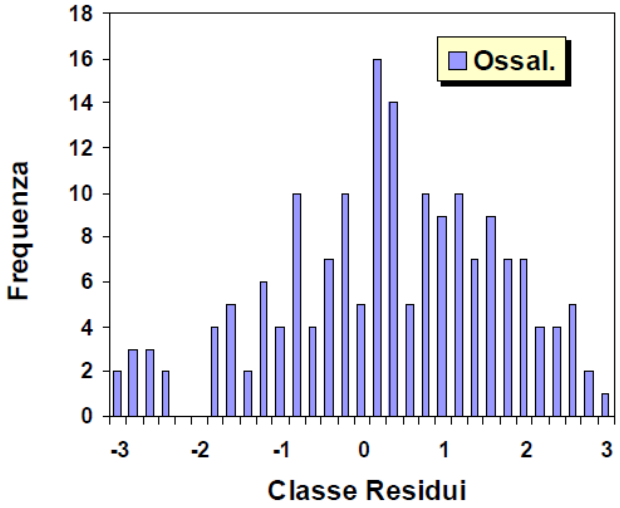
POSITIVE MATRIX FACTORIZATION

Distribuzioni dei residui

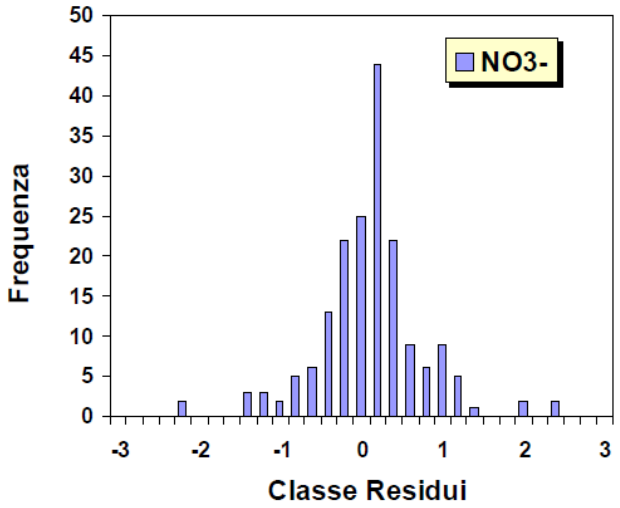
Se le incertezze sono state valutate correttamente e il modello è riuscito a riprodurre bene le concentrazioni di tutte le specie chimiche, le distribuzioni dei residui e_{ij}/s_{ij} dovrebbero essere gaussiane con valori compresi fra ~ ± 2



Se vengono più larghe: incertezza sottostimata oppure è necessario aumentare il n° di fattori



Se vengono molto strette: incertezza sopravvalutata o fattori unici



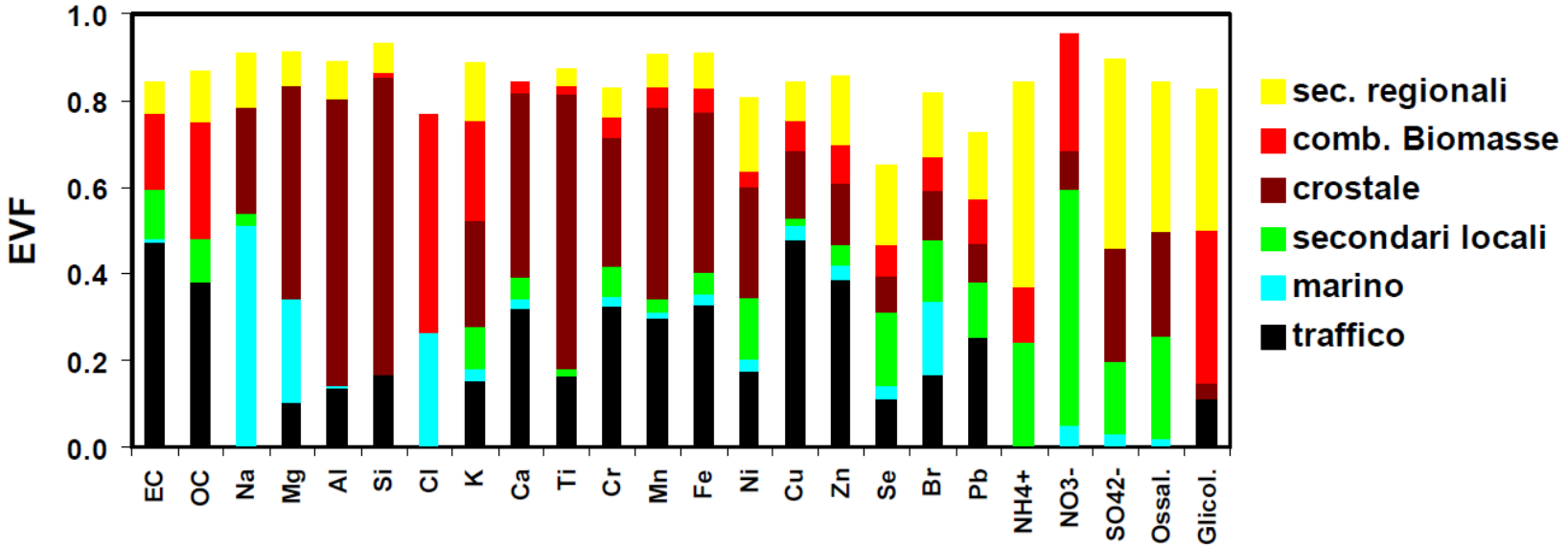
POSITIVE MATRIX FACTORIZATION

$$EVF_{kj} = \frac{\sum_{i=1}^{N} \frac{g_{ik} \cdot f_{kj}}{s_{ij}}}{\sum_{i=1}^{N} \left(\frac{\sum_{k=1}^{P} g_{ik} \cdot f_{kj} + |e_{ij}|}{s_{ij}} \right)} \cong \frac{\sum_{i=1}^{N} g_{ik} \cdot f_{kj}}{\sum_{i=1}^{N} x_{ij}}$$

Explained variations

Contributo medio % della sorgente *K* alla specie *J*

Aiutano a capire l'associazione fattore-sorgente



Se la somma delle *EVF* su tutte le *P* sorgenti ricostruisce male le concentrazioni misurate (< 70%), questo suggerisce di aumentare il numero di fattori

POSITIVE MATRIX FACTORIZATION

Normalizzazione: G ed F assolute

$$x_{ij} \cong \sum_k g_{ik} \cdot f_{kj} = \sum_k (g_{ik} c_k) \cdot \left(\frac{f_{kj}}{c_k} \right) \equiv \sum_k \tilde{g}_{ik} \cdot \tilde{f}_{ki}$$

Output del modello

Scalate per riprodurre pesi e profili fisici: possibile solo conoscendo la massa del PM campionato (come per APCFA)

***G* ed *F* sono determinate a meno di un fattore di scala**

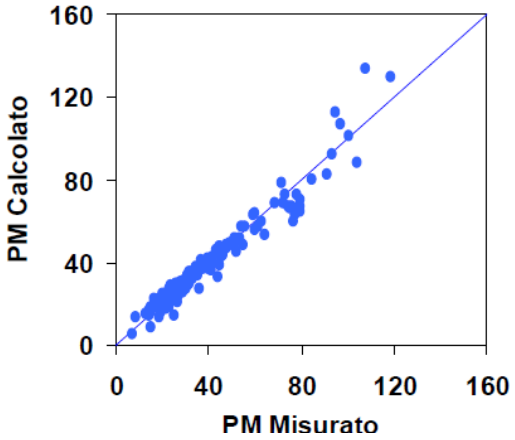
Metodo della massa interna (conc. del PM introdotta come variabile)

$$\tilde{f}_{k,PM} = 1 \Rightarrow c_k = f_{k,PM}$$

Metodo della regressione lineare

$$x_{i,PM} = \sum_{k=1}^{P} g_{ik} c_k + \varepsilon_i$$

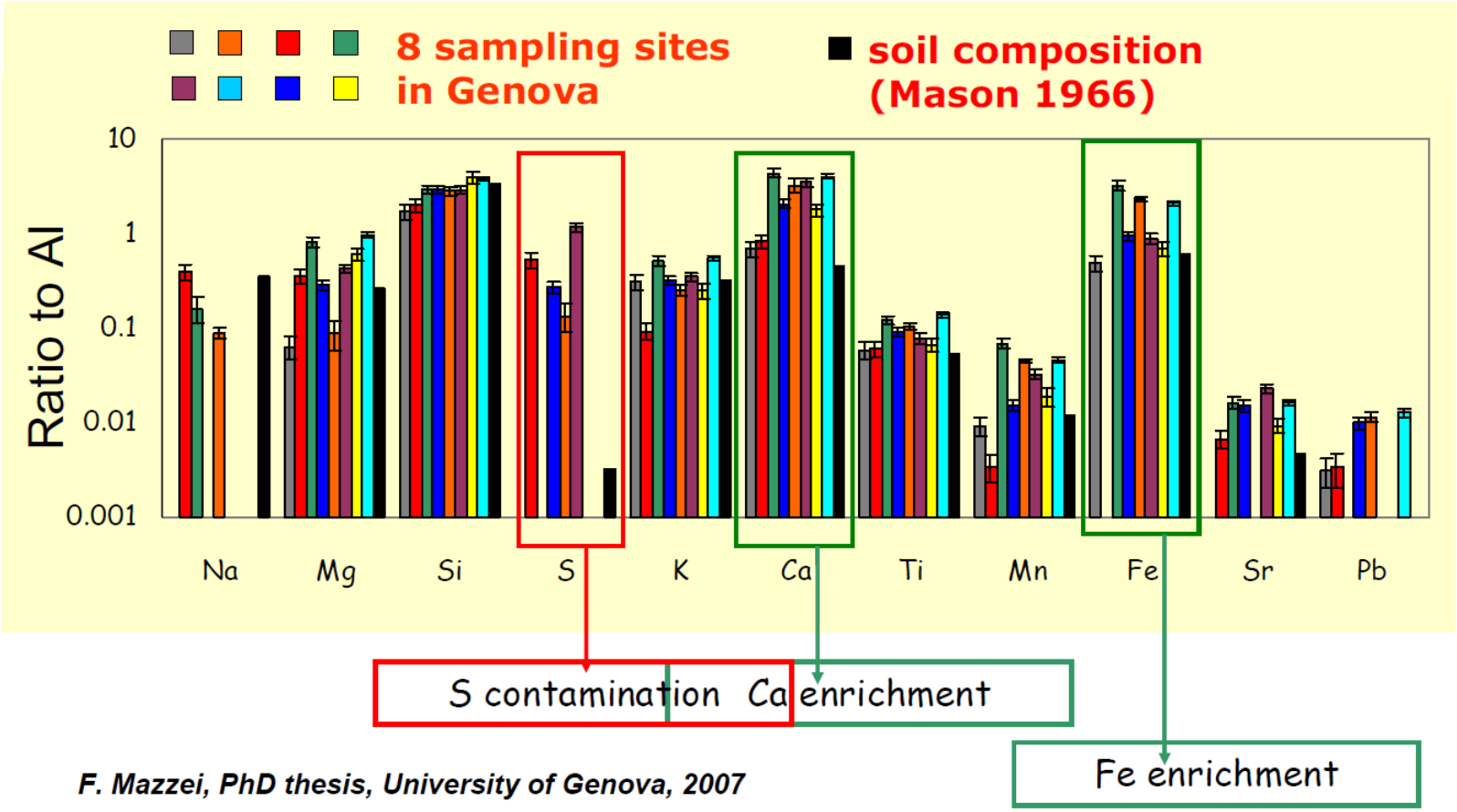
Regressione lineare multipla della concentrazione del PM sui fattori (g_k): i coefficienti della regressione forniscono la normalizzazione



POSITIVE MATRIX FACTORIZATION

Confronto dei profili in diversi siti

Profilo della sorgente CROSTALE



F. Mazzei, PhD thesis, University of Genova, 2007

POSITIVE MATRIX FACTORIZATION

Profili di sorgenti antropiche

TRAFFIC

	Brignole	C. Firenze	Busalla	Multedo
Cu/Pb	4.0 ± 0.7	3.0 ± 0.6	3.2 ± 1.0	3.7 ± 0.6
Cu/Zn	1.0 ± 0.2	0.9 ± 0.2	1.1 ± 0.3	1.1 ± 0.2

Average (PM10)

Cu/Pb = 3.6 ± 0.4

Cu/Zn = 1.0 ± 0.1

HEAVY OIL COMBUSTION

	V/Ni
Multedo	3.8 ± 1.1
Cornigliano 06	2.2 ± 0.8
Lanterna	2.9 ± 0.9
Corso Firenze	3.0 ± 0.7
Via Buozzi	3.1 ± 0.7

Average (PM1, PM2.5, PM10)

V/Ni = 3.1 ± 0.5

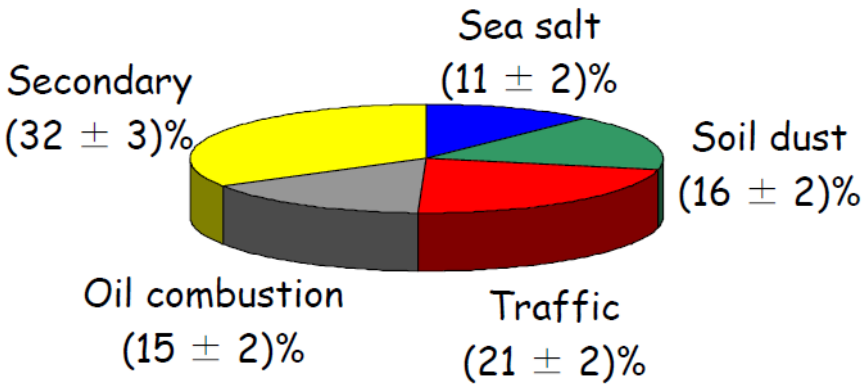
ship exhausts

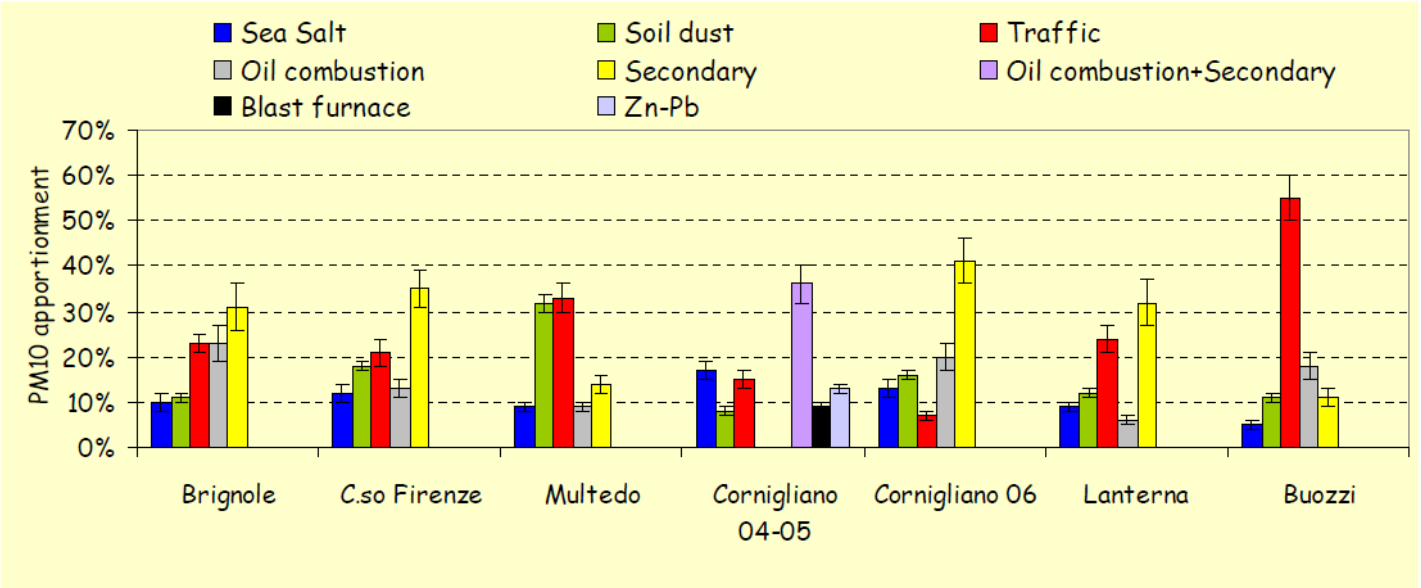
F. Mazzei, PhD thesis, University of Genova, 2007

POSITIVE MATRIX FACTORIZATION

PM10 source apportionment

Average source apportionment in Genova

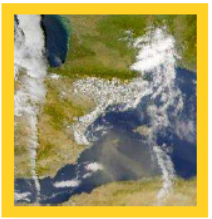


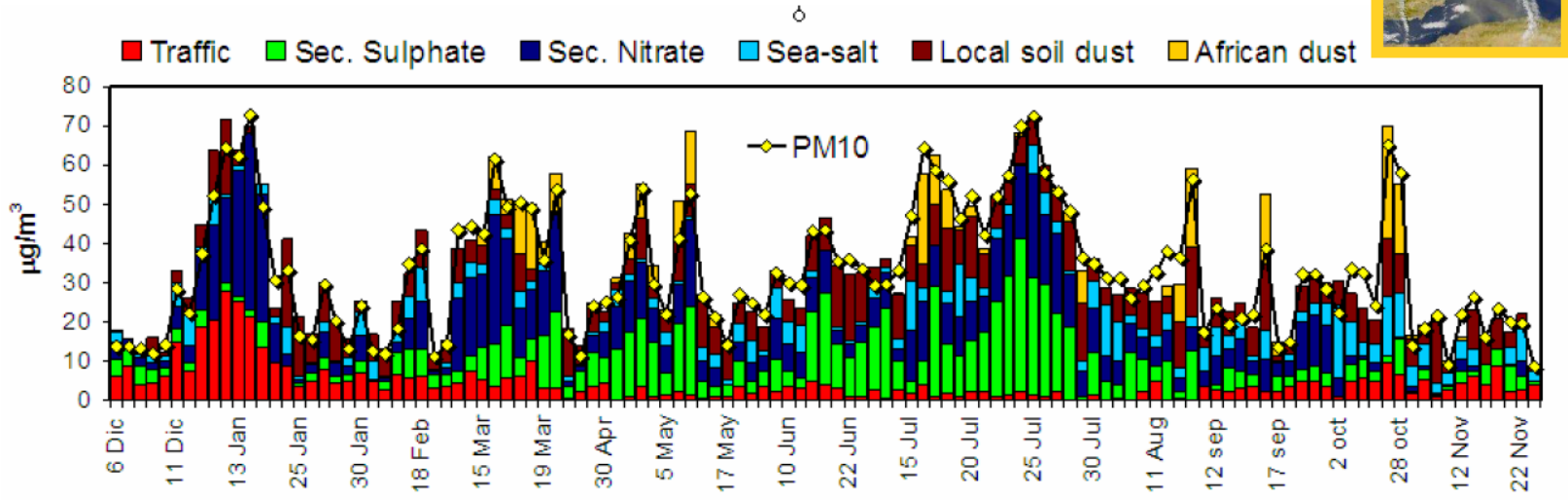


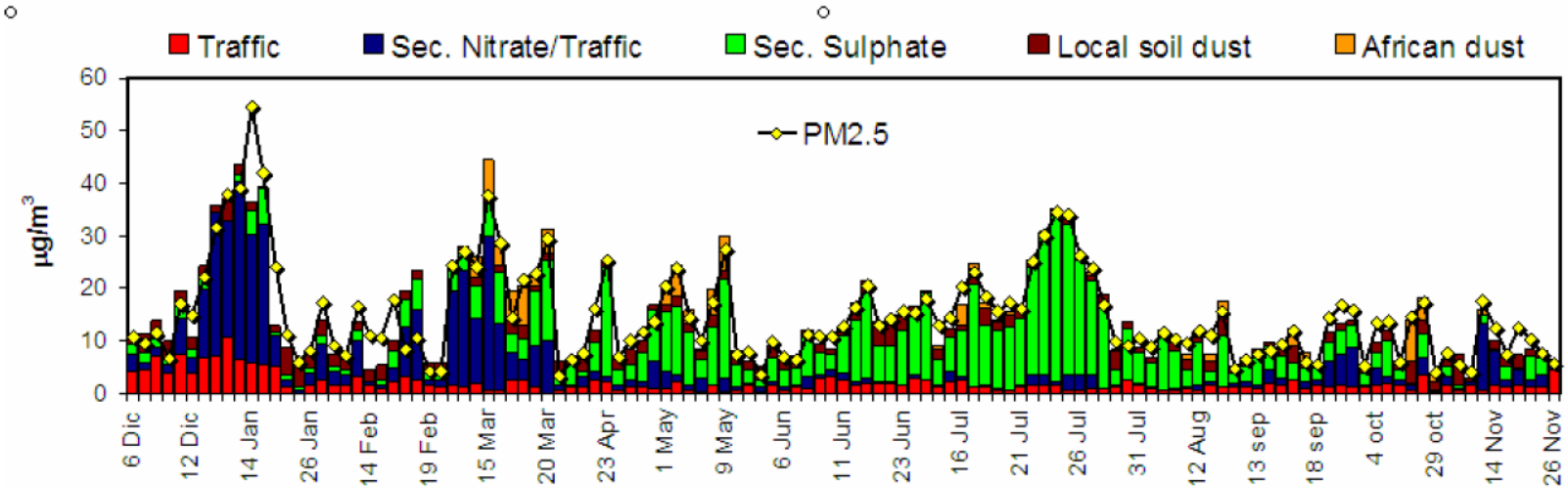
F. Mazzei, PhD thesis, University of Genova, 2007

POSITIVE MATRIX FACTORIZATION

Source apportionment







Progetto APICE (http://www.apice-project.eu/)

Impatto delle emissioni navali sulla qualità dell'aria