Reverse modelling for the determination of fugitive sources of PM10

Guido Cosemans and Clemens Mensink

Integrated Environmental Studies

VITO, Belgium

Reverse modelling using

least squares regression

for the determination of

the source strength of fugitive sources

At problem (X) sites,

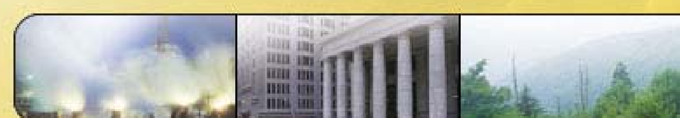
Future air quality standards for a pollutant

may not be respected

with the current source configuration.

Who is responsible?

Emission reduction required ?



Ambient air quality monitoring network (VMM)

(Time series 1 year or more)

- *Measured concentrations of pollutant*
    - *at problem site*
    - *at remote 'clean air sites' (background)*
- *Synchronously measured meteorological data*

Averaging time of the measurements

Meteorological parameters:

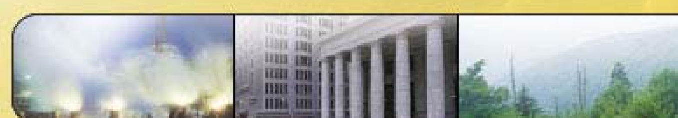
½hourly concentrations measured

PM10:

½hourly concentrations measured

Heavy metals, PAH ... :

24h-average concentrations measured



Time series (TS) of concentrations at X-site

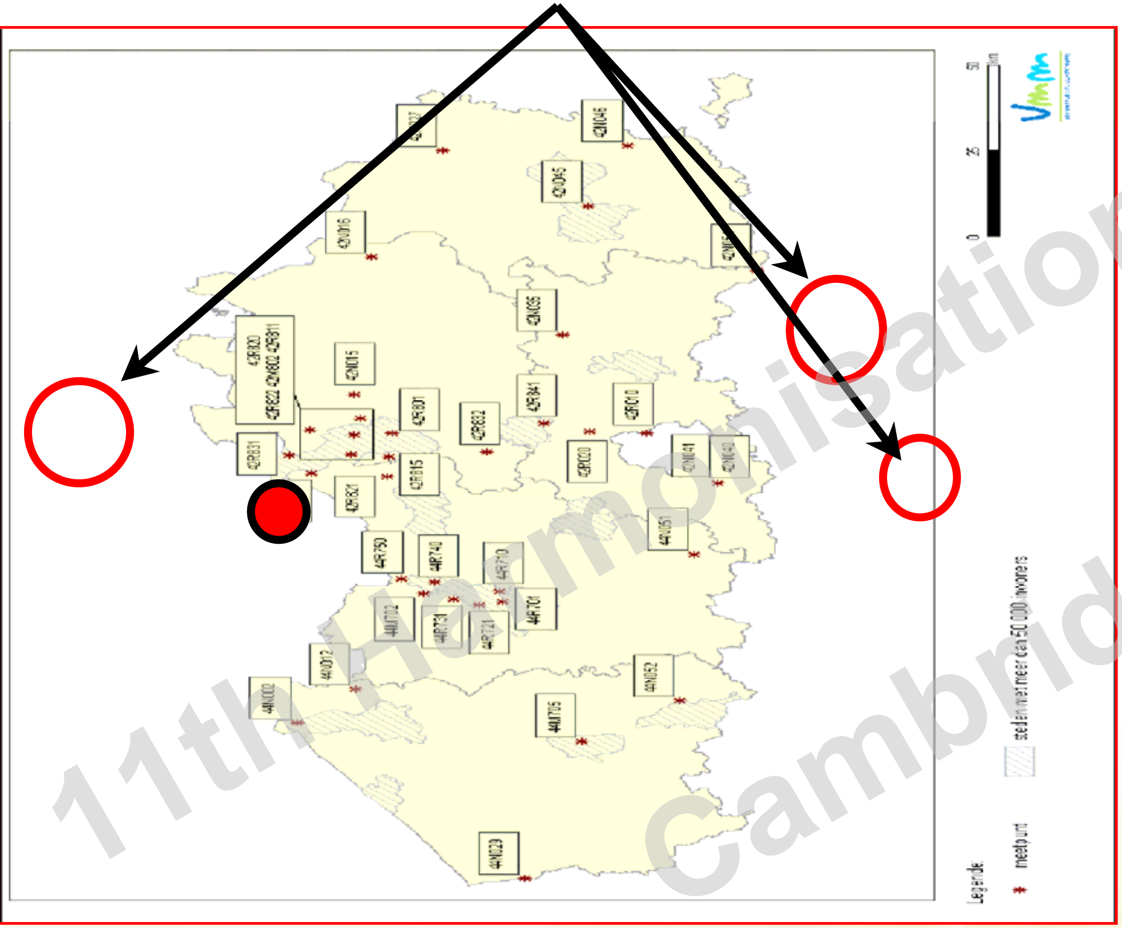
Cumulative Frequency Distribution (Diagram on Log-Prob.paper)

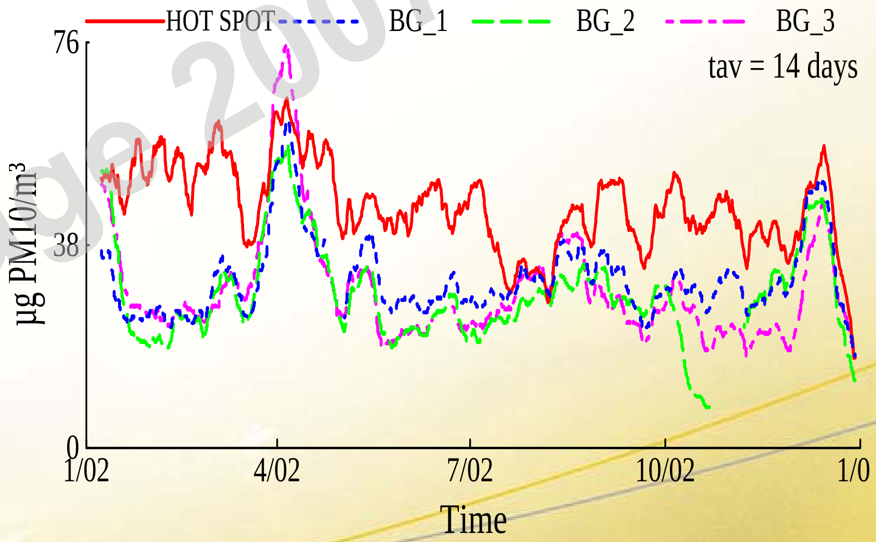
TS of regional background (RBG)

TS(X) – TS(RBG) → TS(impact local sources at X)

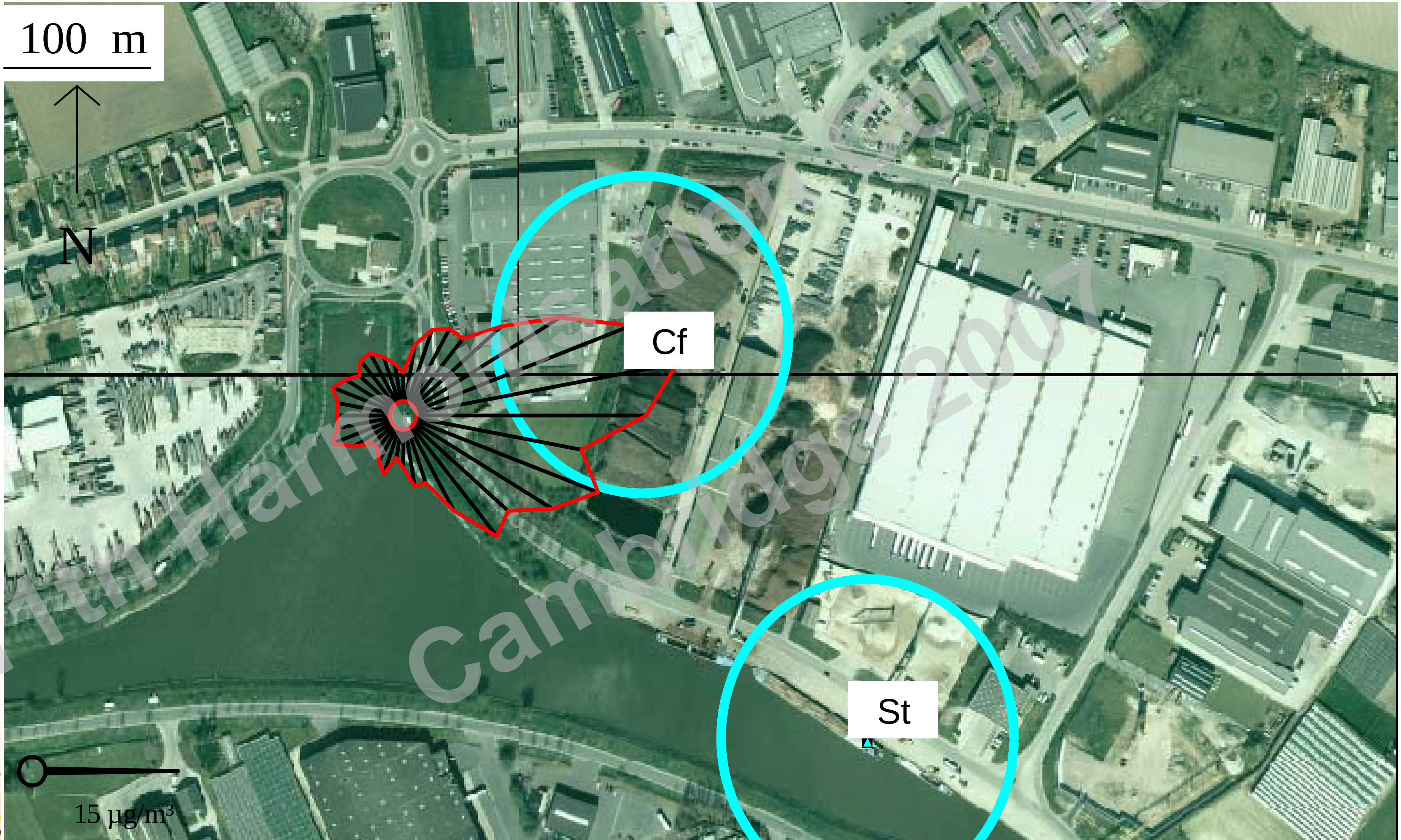
Pollutant roses

On *TS(X) – TS(RG) -> TS(impact local sources at X):*





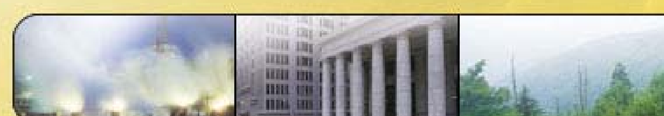
PM_{10} monitoring site X, Compost facility, Sand trader



Lessons from pollutant rose

Wind sector	average concentration (µg PM10/m³)	% of time	% of total concentration
50°-150°	21.5	19	48
other	5.3	81	52
all	8.4	100	100

- 8.4 µg/m³ above regional average
- <u>48 % due to important local sources</u>
- 52% due to uniformly distributed small town PM10 sources



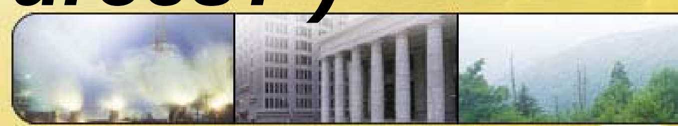
For future air quality standards to be respected :

Where are the sources?

Emission reduction required ?

(Authorities: impact to be reduced by 75 %)

(Plant operator: What sources?)



Investigate variation of concentrations with:

Wind speed

Time (hour of day, day of week, season...)

- per wind sector -

using only measured time series;

comparing model prediction with observations

Reverse modelling using

least squares regression

for the determination of

the source strength of fugitive sources

The system of equations

For every ½hour or day j :

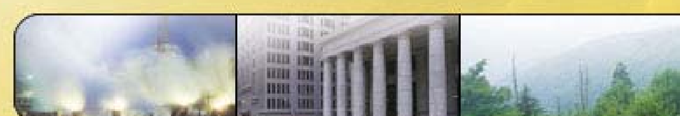
$$\mathbf{Observed}_j = \Sigma_{i=1,NS}\, \delta_{ij}^{\;-1}\, x_i \quad \text{(= Prediction)}$$

$\delta_{ij}^{\;-1}$: impact unit emission at place i and time j

(if t_{av}=1 day, $\Sigma_{k,\; k=1,24}$ hours)

x_i : unknown source term,

to be solved by regression



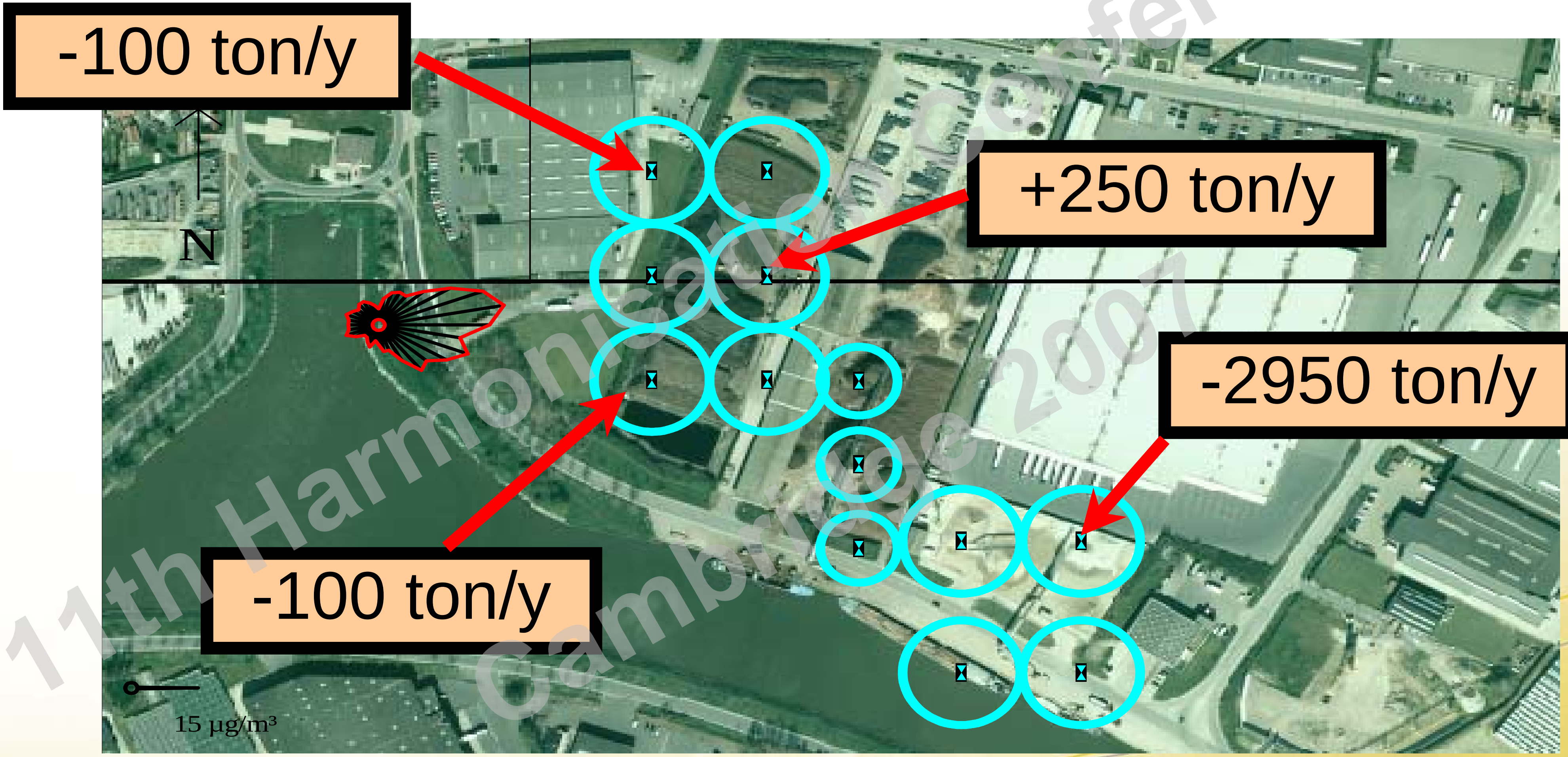
$N_{Sources}$ = 13 unknown

J equations:

j= 1,365 (day)

j=1,17528 (½hour)

Typical regression solution:



What happened ? What does this mean?

Least Squares Regression

would like to express its gratitude

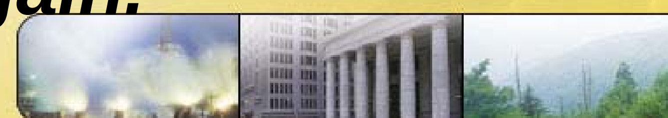
for you to have supplied it

with so many opportunities

<u>for noise fitting.</u>

Please enjoy this solution with standard deviations of coefficients up to 5000 %

We hope to see you again.



SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.12
R Square	0.02
Adjusted R Square	-0.18
Standard Error	3.06
Observations	7.00

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1.00	0.73	0.73	0.08	0.79
Residual	5.00	46.70	9.34		
Total	6.00	47.43			

	Coeffi-cients	***Standard Error***	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	4.56	**4.71**	0.97	0.38	-7.54	16.65
x	-0.18	**0.65**	-0.28	0.79	-1.86	1.49

Answers you should know:

How to avoid noise fitting ?

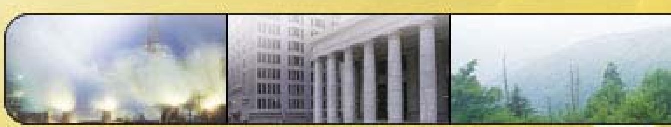
By using few sources

What is the source of noise fitting?

The very nature of Least Squares Regression

What is the very nature of Least Squares Regression?

It is to minimize the sums of squares of differences



The very nature of Least Squares Regression

... is to minimize the sum of squares of differences between observations C_j and predictions $\Sigma_{i=1...NS}\ \delta_{ij}^{\ -1}\ x_i$:

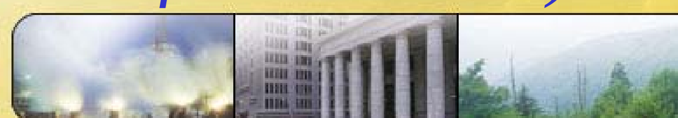
$$\boldsymbol{\Sigma_{j=1...NDAYS}\ \{C_j - \Sigma_{i=1...NS}\ \delta_{ij}^{\ -1}\ x_i\}^2}$$

Minimum *not* for **x_i = source strengths Q_i** at *i*

but, *(noise fitting)*

for $x_i = Q_i \pm \Delta_i$, whith $\Delta_i \gg Q_i$

(Δ_i is the noise fitting component of the solution)



Noise fitting

- why it happens
- how to avoid it
- how to recognize it (***standard deviations of coefficients)***

Literature

- ***ill-conditioned system***
- ***collinearity***

Other aspects of regression: wind direction-t_{av} (1/2)

- Uncertainty on wind direction

- measured wind direction ↔

  transport wind direction

(See also: tracer gas experiments:

not measured wind direction,

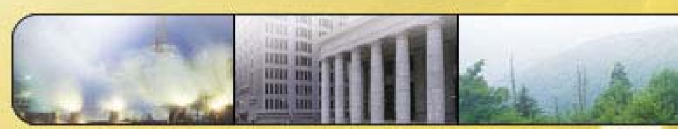
but maximum in cross wind concentration profile)

Other aspects of regression: wind direction-t_{av} (2/2)

- Regression on day averages explains greater part of observations than regression on ½hourly concentrations .

  - t_{av}= ½hour: difference in ½hourly transport/measured wind direction:
    - observation can not be explained by LSQ.

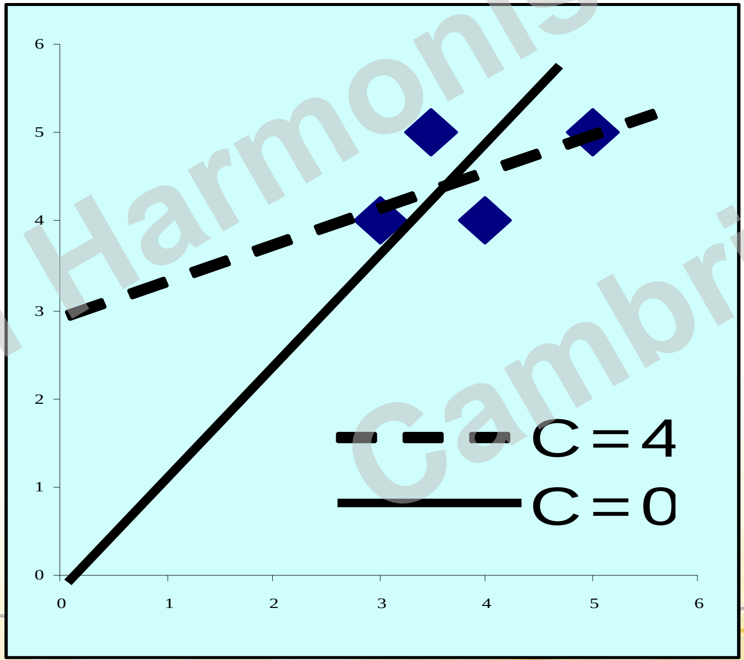
  - t_{av} = 1 day: such differences only important if they lead to a different frequency distribution of wind over 24 h.



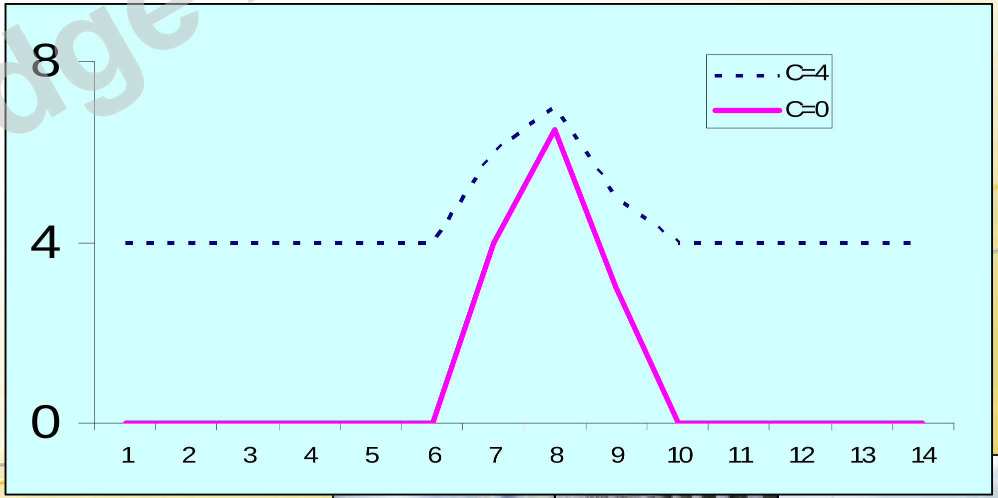
Other aspects: regression with/without constant (1/2)

- Regression with constant → lower value for source term.
- Can be graphically explained:

y=f(x)



observed (y) versus predicted (x) time series



Other aspects: regression with/without constant (2/2)

Regression can NOT take the effects of

random emissions

into account.

(Solution:

- group hours with such events in a separate system of equations;
- proper interpretation of regression constant (?)

)

Other aspects of regression: correlation coefficient (1/2)

As LQR uses noise fitting to obtain the ‘best’ solution,
-having the highest correlation- ,

the correlation coefficient between
observation and prediction
is not very useful -

for as far the observation and prediction
are evaluated over the same
averaging time and observation period
as used to build the system of equations.

Other aspects of regression: correlation coefficient (2/2)

Alternative evaluation criteria:

pollutant roses

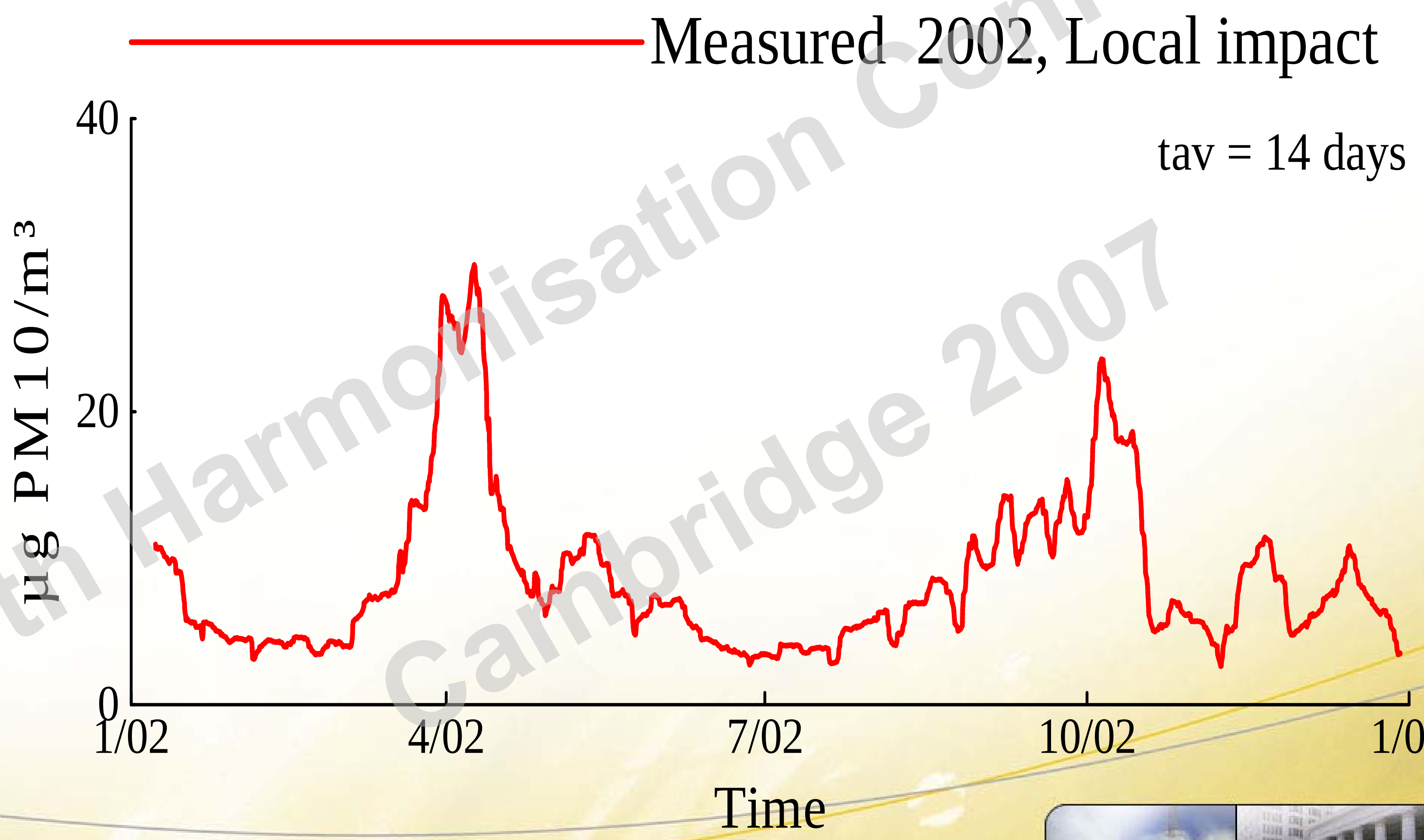
time series over other averaging time

time series for an other period

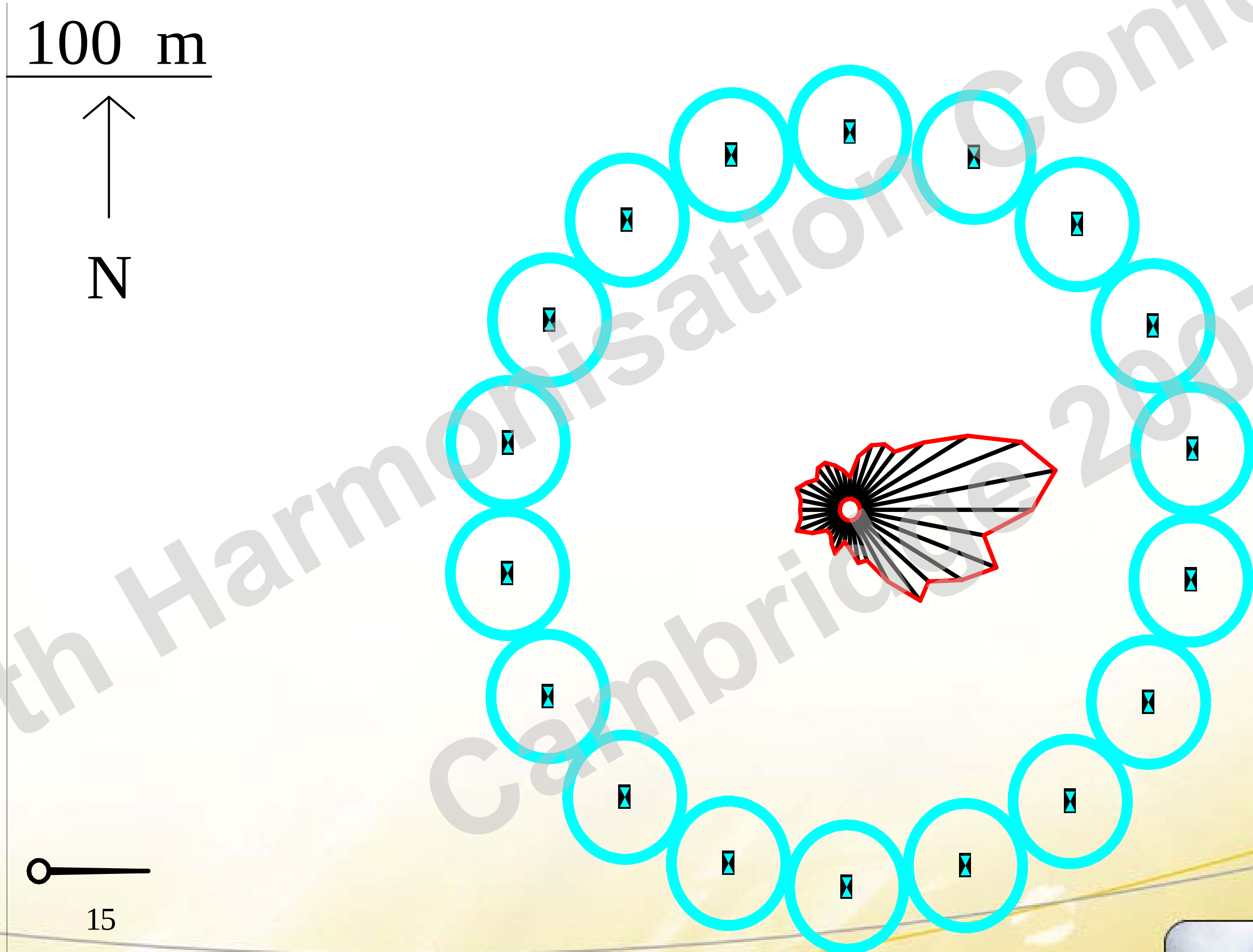
Idem for:

- cumulative frequency distributions

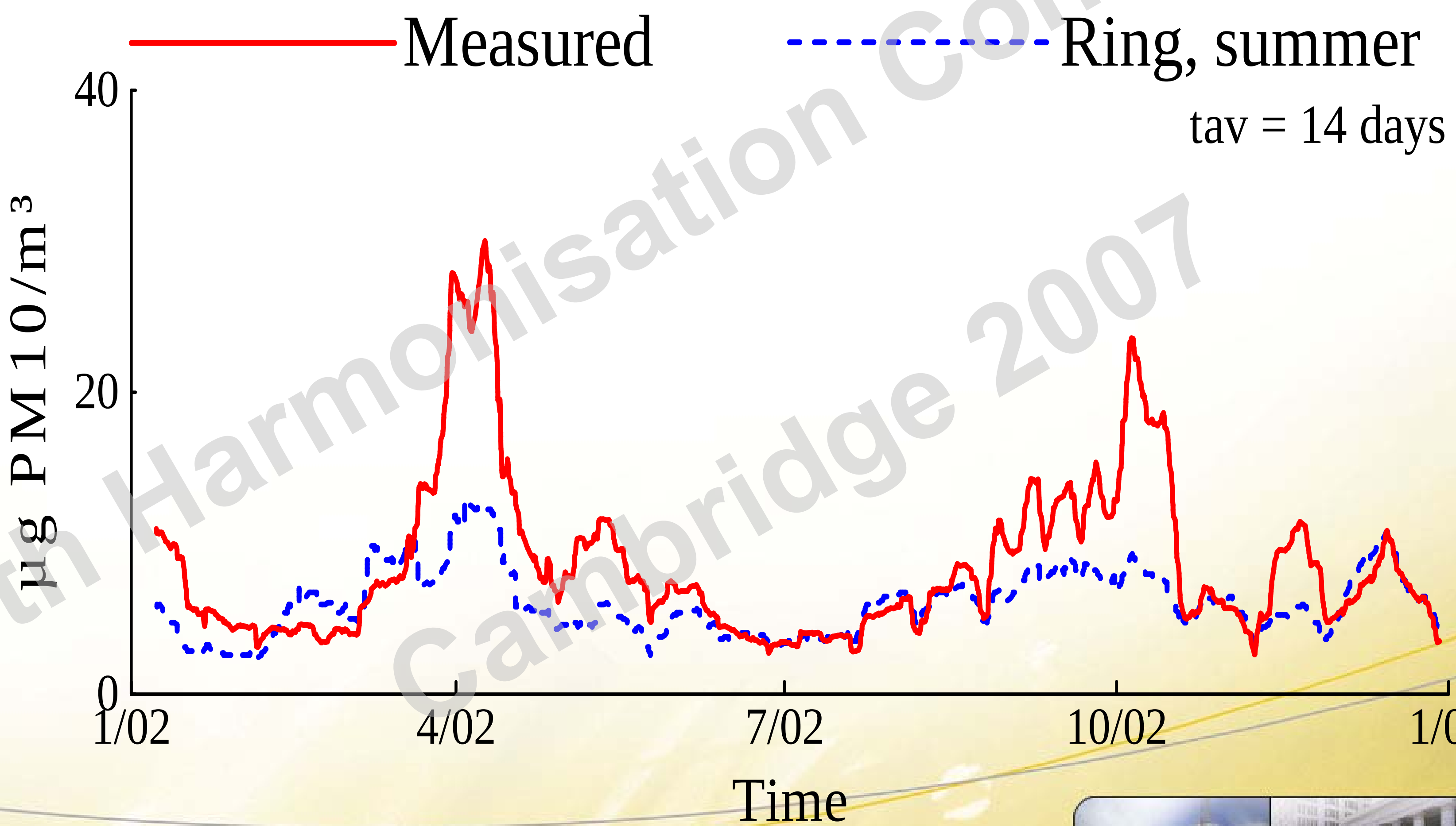
Time series local impact



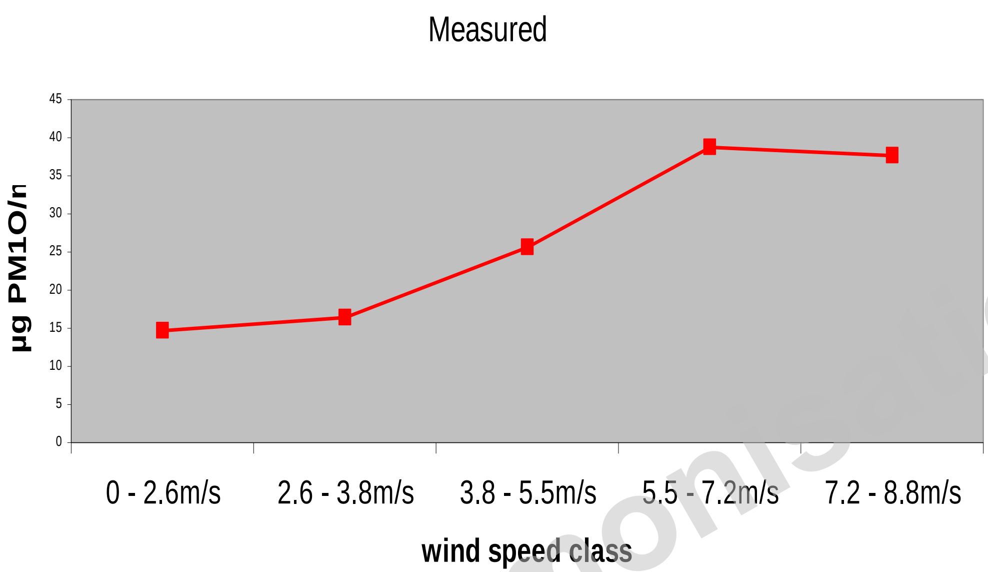
Source configuration for regression

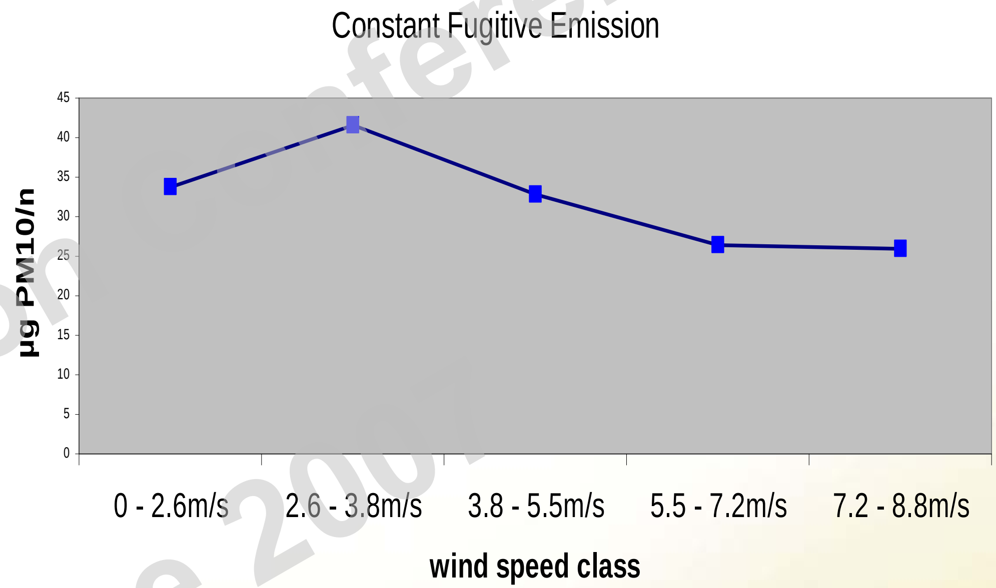


Time series predicted by LSQ solution



Wind speed dependency of observed and of modelled concentrations do not agree



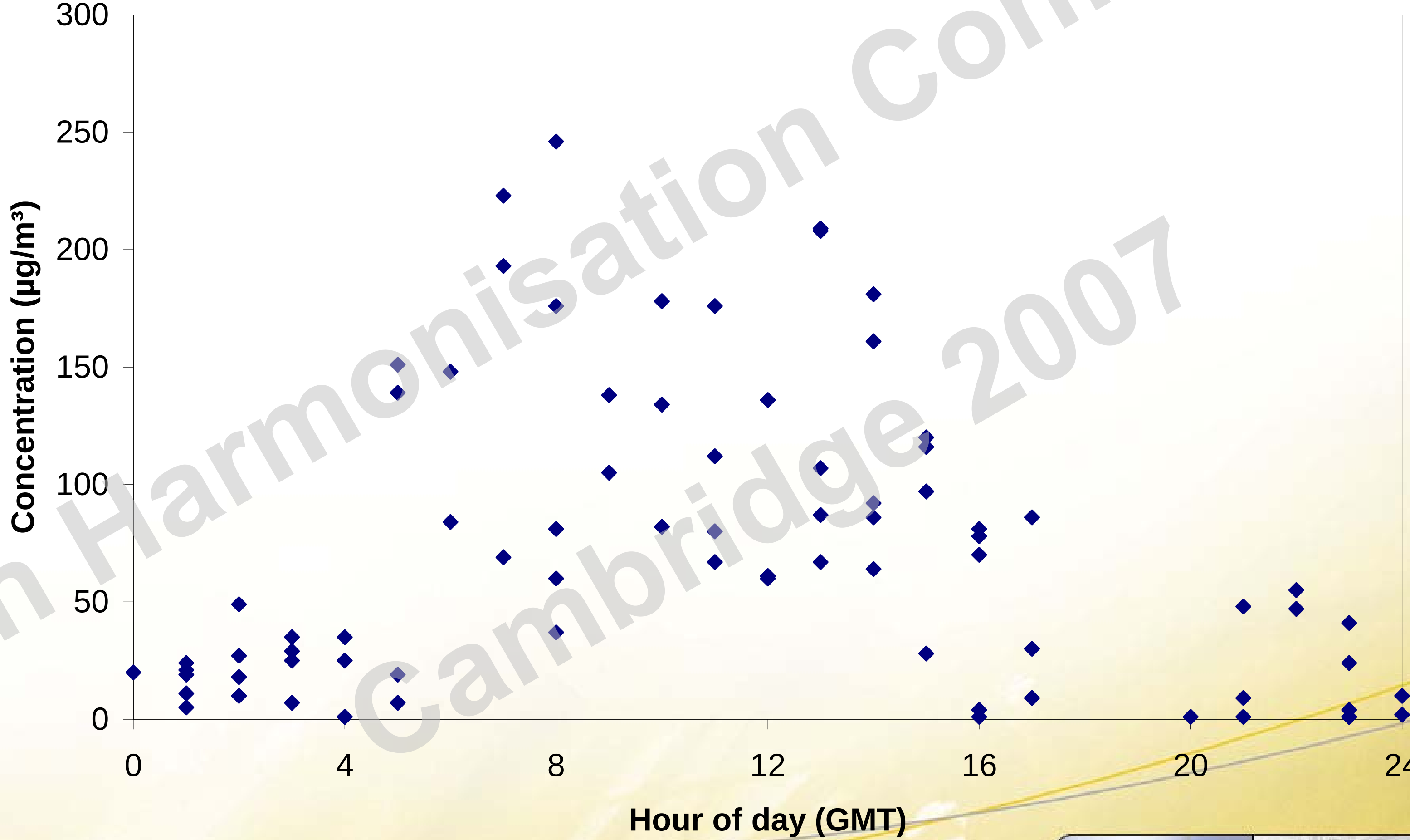


Measured: increases with increasing wind speed

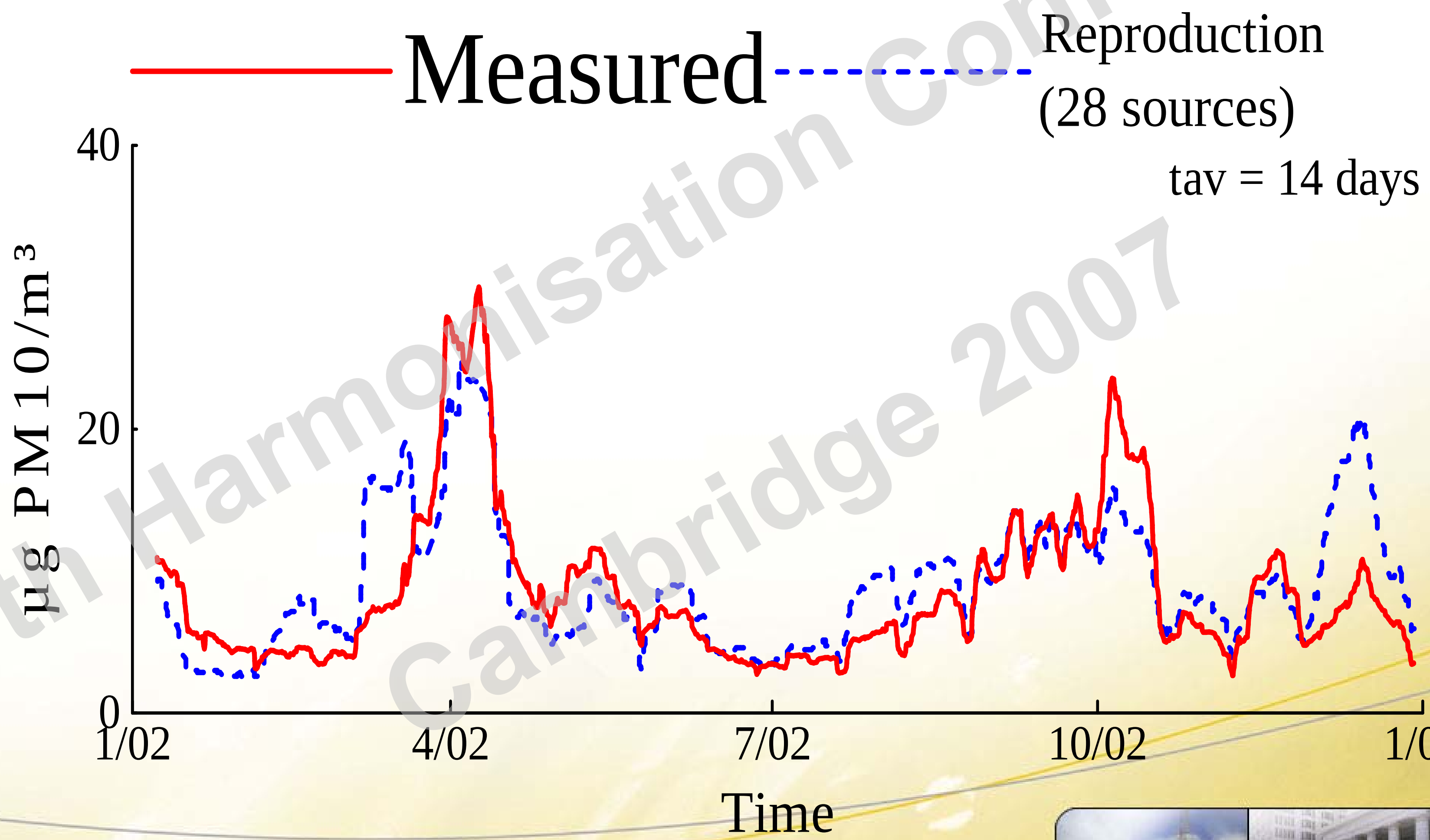
Calculated (constant Fug. sources) : decreases with increasing wind speed

*'unit emission' = {1 times max(0,min(8,(**u**-2)))³} tons/year*

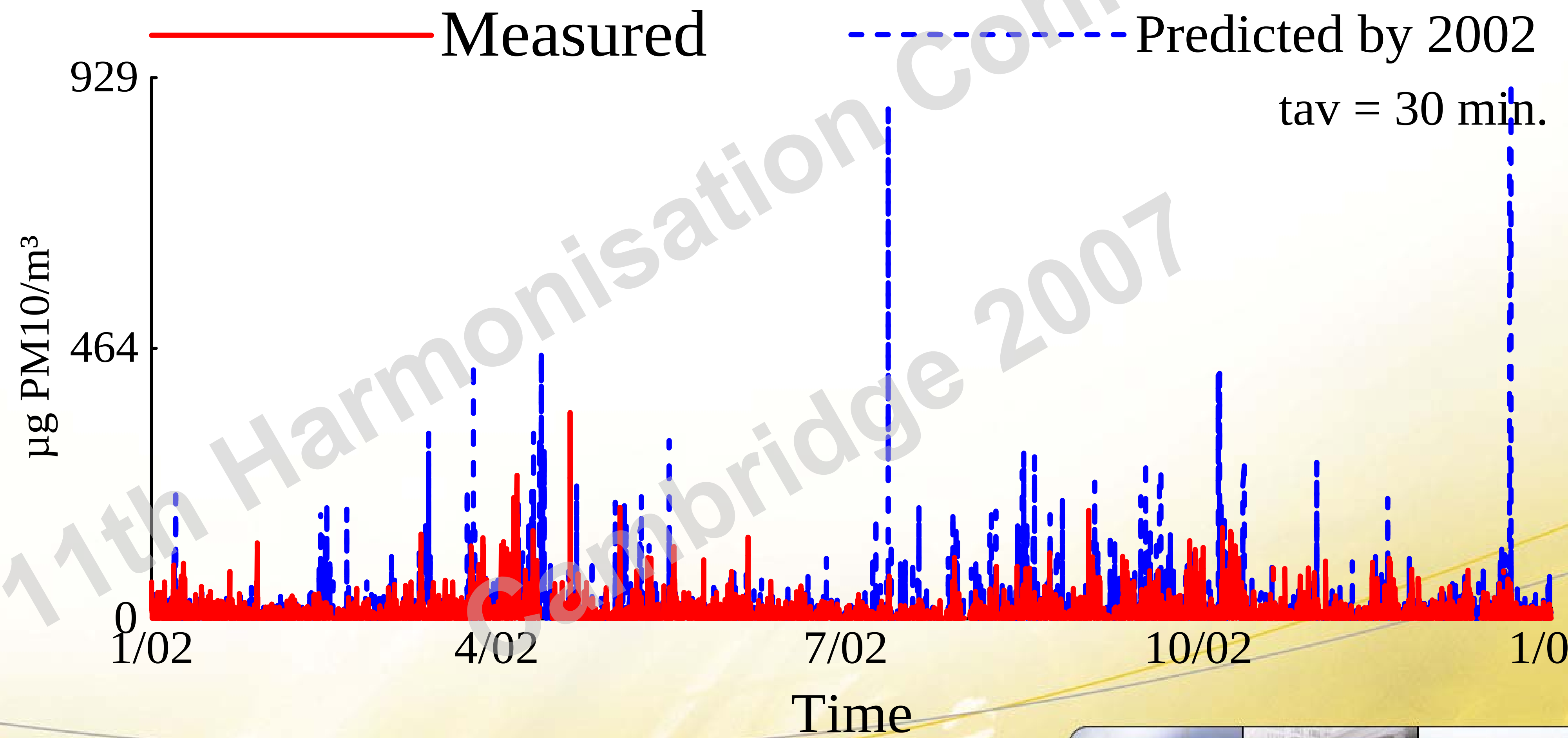
Observed local impact per hour during April peak



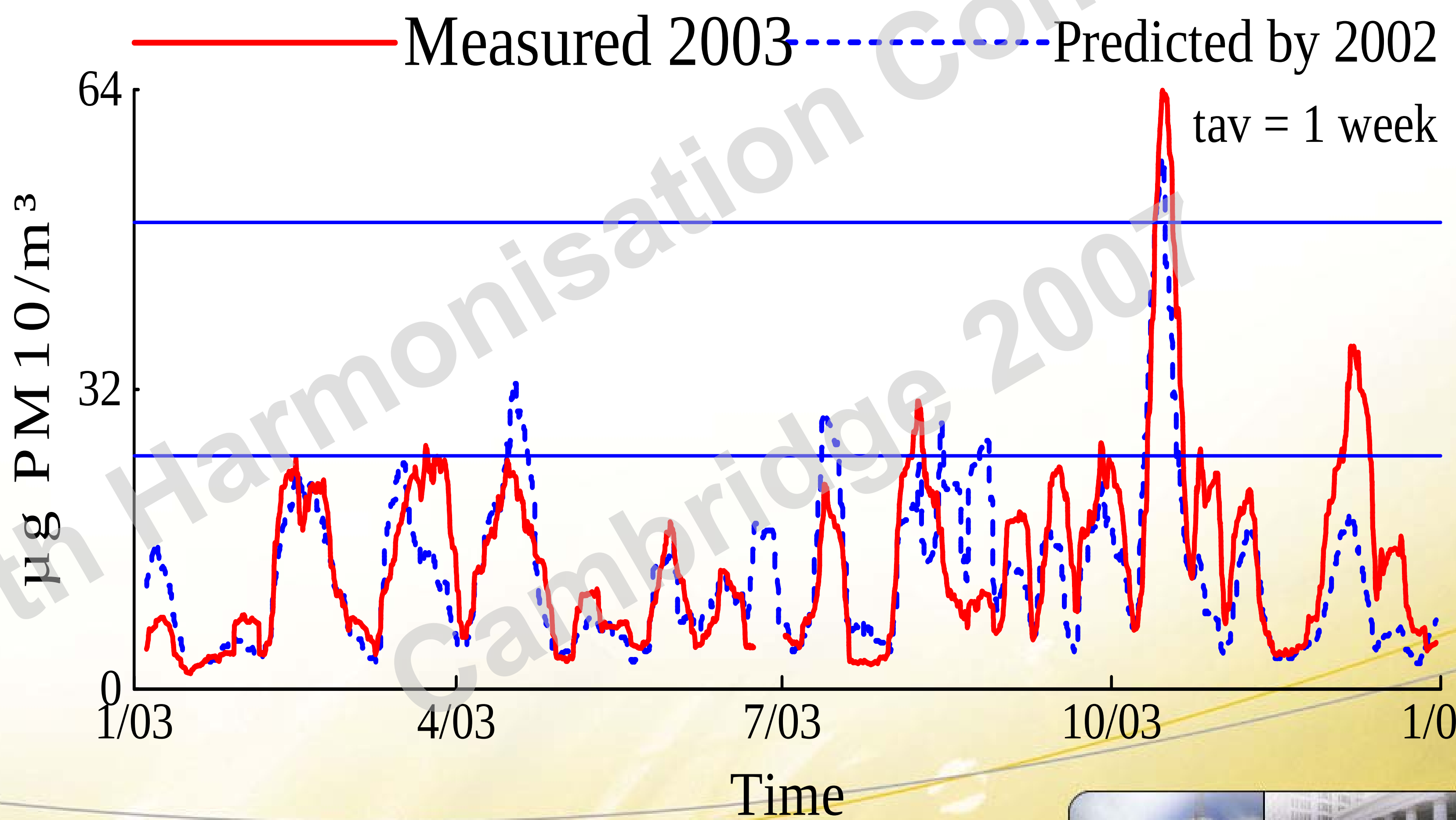
Regression on October peak explains April peak ($t_{av\ 1\ day}$) (Year 2002)



Regression on October produces too high peaks for t_{av} ½ hour (Year 2002)



Correction: Weaken responsible source
... and so on, till all criteria OK. Then: evaluation on 2003



Conclusions

- **Reverse modelling leads to a better understanding of the observed time series;**
- **For as well ½hourly data as 24h data;**
- **No emission factors needed;**
- **Time and wind speed dependencies tell more on the nature of the fugitive source;**
- **Regression must be kept under straigth control to avoid noise fitting;**

Action 2: Start with 1 unknown diffusive source

System of equations, one equation per DAY:

$$\left(\sum_{h=1}^{48} C_j\right) - \sum_{k=1}^{NP}\left(\sum_{h=1}^{48} P_{kh}\right) = \sum_{i=1}^{NFS}\left(\sum_{h=1}^{48} \partial_{ih}^{-1}\mathbf{Q}_{\boldsymbol{i}}\right)$$

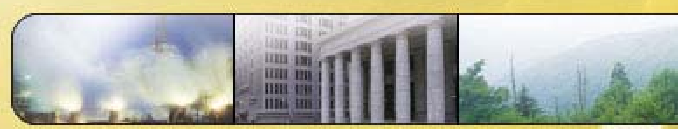
sum of measured ½hourly concentrations over day

computed impact of NP known (point) sources

impact of NFS diffusive sources:

$\mathbf{Q}_{\boldsymbol{i}}$: unknown source strength;

δ_{ih}^{-1} : impact of unit emission



Impact of unit emission (1/2)

$$\partial_{ih}^{-1}$$

impact of unit emission for unknown source i at hour h

unit emission: 1 mass unit per time unit, BUT

- having the same time dependency as the real source.

(e.g.: if only active during working days from 8 am till 18 pm, then unit emission is 1 during these hours and 0 otherwise.)