INAIL
ISTITUTO NAZIONALE PER L'ASSICURAZIONE CONTRO GLI INFORTUNI SUL LAVORO

USE OF NEURAL NET MODEL TO FORECAST CARDIOVASCULAR AND RESPIRATORY DISEASES FROM METEOROLOGICAL AND POLLUTION DATA.

*Armando Pelliccioni(*) and **Rossana Cotroneo(**)**,*
(*) Ispesl-Dipia, Via Fontana Candida 1, 00040 Monteporzio Catone, Roma
(**) Istat, Viale Leigi, 13 00189 Roma

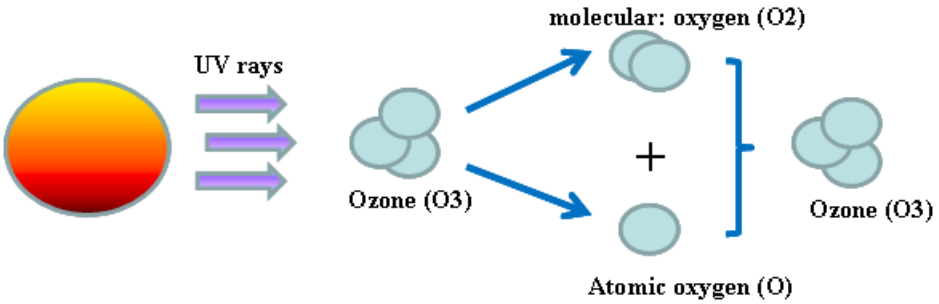
Kos Island, 2-6 October, 2011

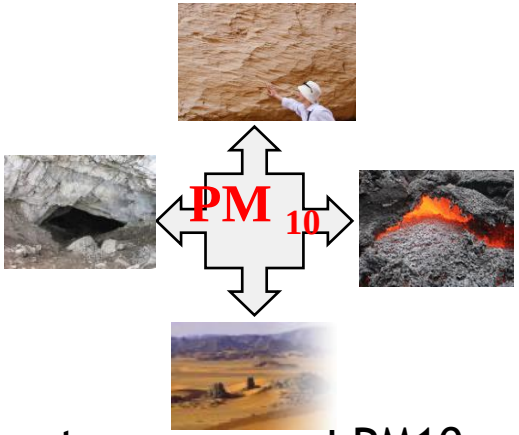
CONTENTS

Introduction

Air quality

- The degradation of air quality is produced by high levels of ozone (O_3) and PM10





- We analysed exposure to outdoor pollution due to ozone and PM10
- We apply Neural Network to analyze the annual trend of morbidity and mortality in hospital for cardio-respiratory diseases due to PM10 and O3
- The NN was trained to predict up to five days in advance the number of mortality in hospital for cardio-respiratory diseases per day.

Dataset description

INAIL
ISTITUTO NAZIONALE PER L'ASSICURAZIONE CONTRO GLI INFORTUNI SUL LAVORO

Environmental Time Series

1. Carbon monoxide (mgm-3) – CO-
2. Nitrogen Oxide (µgm-3) – NO-
3. Nitrogen Dioxide (µgm-3) - NO2-
4. Mono-Nitrogen Oxides (µgm-3) -NOx-
5. Ozone (µgm-3) - O3 -
6. PM10 (µgm-3) - PM10 -
7. Temperature (C) - Temp
8. Relative humidity (%)- RH
9. Pressure (mbar)- Press
10. Rain (mm)
11. Global solar radiation (W\m^2) -GSR
12. Hour
13. Julian day

Hospital Discharge Records

1. Address
2. Age
3. Sex;
4. Admission and dismissal Date
5. Type of care
6. Admission type
7. Number of days in hospital
8. Source of admission
9. Patient's Status,
10. Principal Diagnosis Codes (ICD9CM)
11. Mode of arrival at hospital

Source: *Database Brace (ISPRA)*

Source: *Database EPICS*

Standardize data

$$z_{if} = \frac{x_{if} - m_f}{s_f}$$

Department of Production Plants and Human Settlements

General statistics

2005		Mean	Sd	Skewness	Kurtosis	N	Missing	Min	Max	CV (%)
CO	(mg/m³)	1.06	0.82	2.99	16.70	40795	3005	0.00	12.16	0.77
NO2	(µg/m³)	63.42	31.85	0.53	0.26	40773	3027	0.45	249.84	0.50
NO	(µg/m³)	47.85	67.62	2.95	13.11	40784	3016	0.00	860.07	1.41
Nox	(µg/m³)	136.76	126.82	2.40	9.41	40806	2994	0.50	1365.47	0.93
O3	(µg/m³)	35.99	35.30	1.12	0.75	24300	19500	0.00	220.54	0.98
PM10	(µg/m³)	36.33	17.15	1.52	5.06	23478	20322	0.00	202.49	0.47
Rain	(mm)	0.09	0.64	14.19	290.46	8494	35306	0.00	20.00	6.88
Press	(mbar)	1016.14	6.80	-0.42	0.95	17396	26404	986.47	1035.81	0.01
GSR	(W/m2)	133.56	230.59	1.88	2.44	34394	9406	0.00	1031.11	1.73
Temp	(C°)	14.37	8.12	0.10	-0.70	43145	655	-5.90	41.61	0.57
RH	(%)	70.21	19.49	-0.32	-0.83	43143	657	0.05	99.14	0.28
Cardiac Admission		172.48	50.30	-0.60	-.29	365	0	3	286	29.16
Respiratory Admission		78.74	25.51	0.07	-0.14	365	0	1	149	32.40

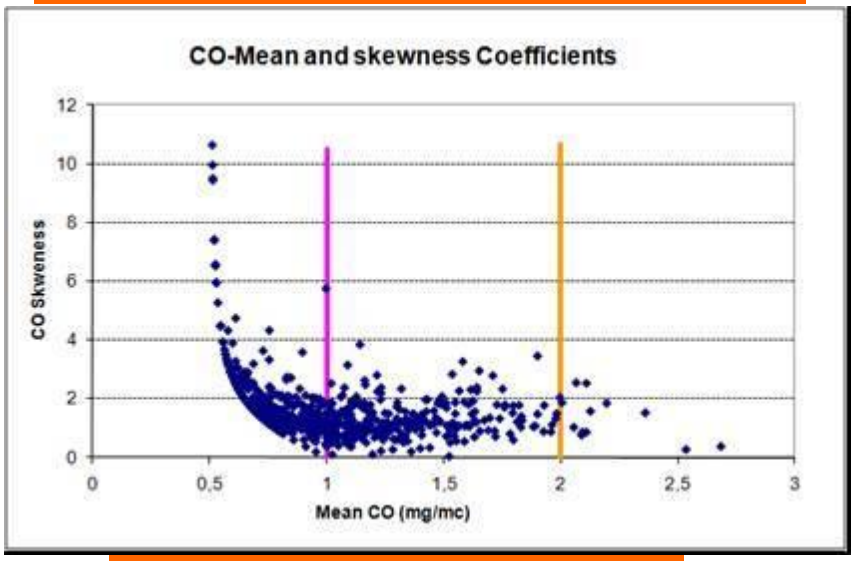
2006		Mean	Sd	Skewness	Kurtosis	N	Missing	Min	Max	CV (%)
CO	(mg/m³)	0.99	0.75	2.68	12.46	36322	7478	0.00	11.34	0.75
NO2	(µg/m³)	69.80	35.47	0.49	0.15	40385	3415	0.00	250.19	0.51
NO	(µg/m³)	50.50	71.78	2.85	12.95	40392	3408	0.00	1228.35	1.42
Nox	(µg/m³)	147.18	134.82	2.25	8.17	40400	3400	0.00	1734.78	0.92
O3	(µg/m³)	36.63	35.41	1.04	0.63	15248	28552	0.00	227.59	0.97
PM10	(µg/m³)	39.83	17.35	0.80	0.35	33102	10698	0.18	104.90	0.44
Rain	(mm)	0.60	5.53	18.11	435.33	8697	35103	0.00	182.00	9.17
Press	(mbar)	1017.51	6.06	-0.07	0.45	17401	26399	996.00	1040.00	0.01
GSR	(W/m2)	138.38	230.18	1.78	2.08	34853	8947	0.00	1005.00	1.66
Temp	(C°)	15.64	7.79	0.26	-0.56	42813	987	0.00	44.00	0.50
RH	(%)	70.25	19.92	-0.37	-0.79	34794	9006	1.00	99.00	0.28
Cardiac Admission		167.78	48.21	-0.60	-0.31	364	1	7	259	28.73
Respiratory Admission		72.89	21.09	-0.08	-0.04	364	1	6	134	28.93

Time series

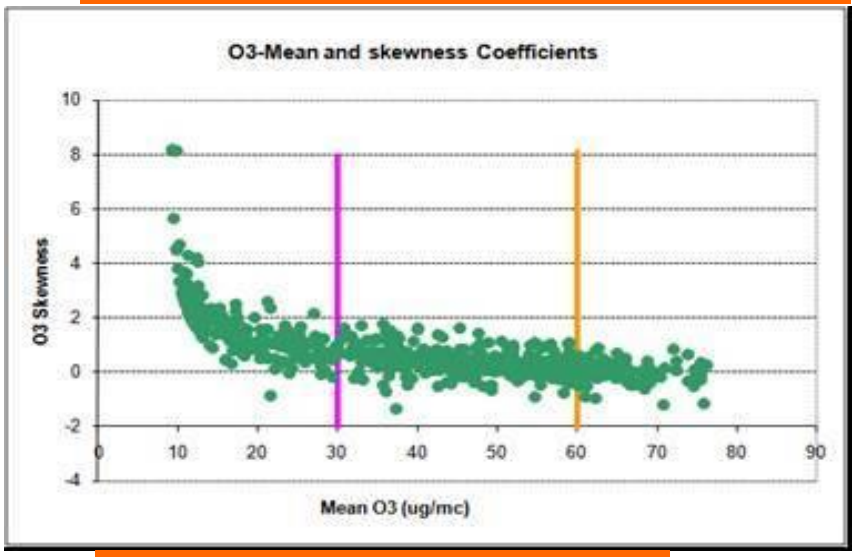
Skewness

ISTITUTO NAZIONALE PER L'ASSICURAZIONE CONTRO GLI INFORTUNI SUL LAVORO

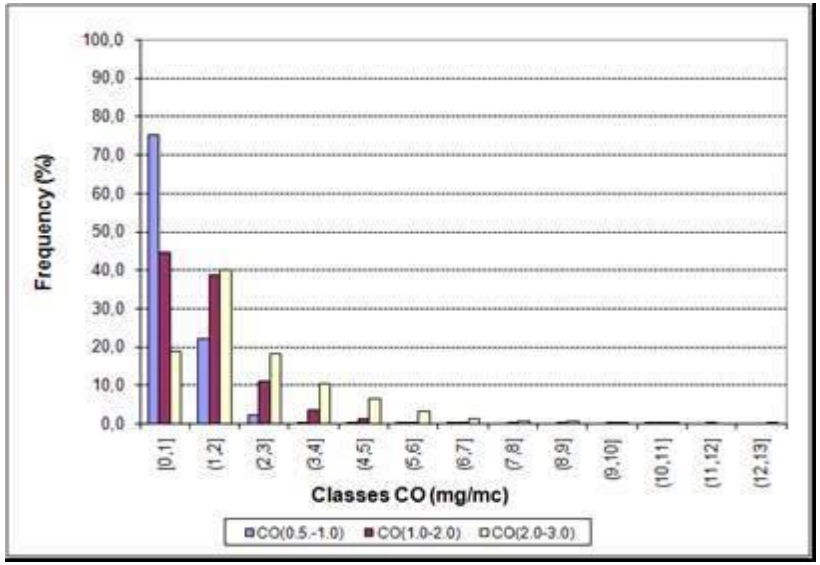
Relation between Mean(CO) and Skenewss(CO)



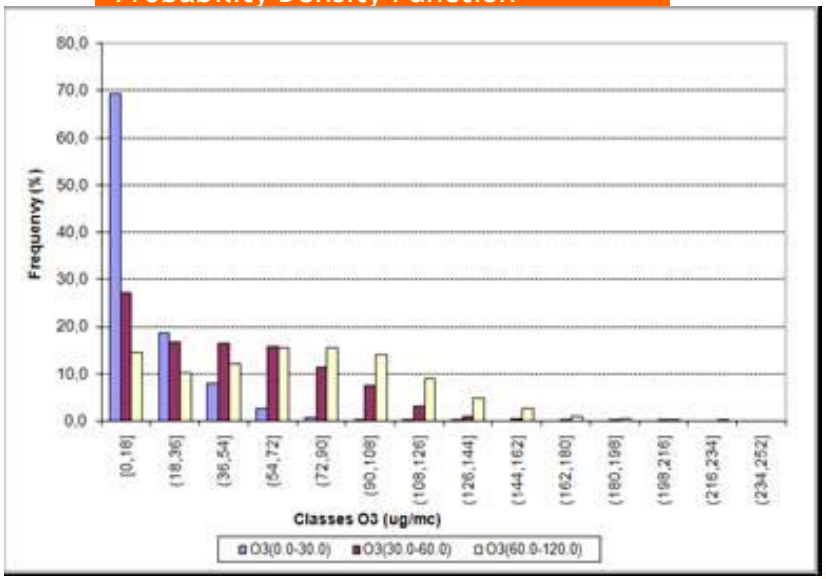
Relation between Mean(O3) and Skenewss(O3)



Probability Density Function



Probability Density Function

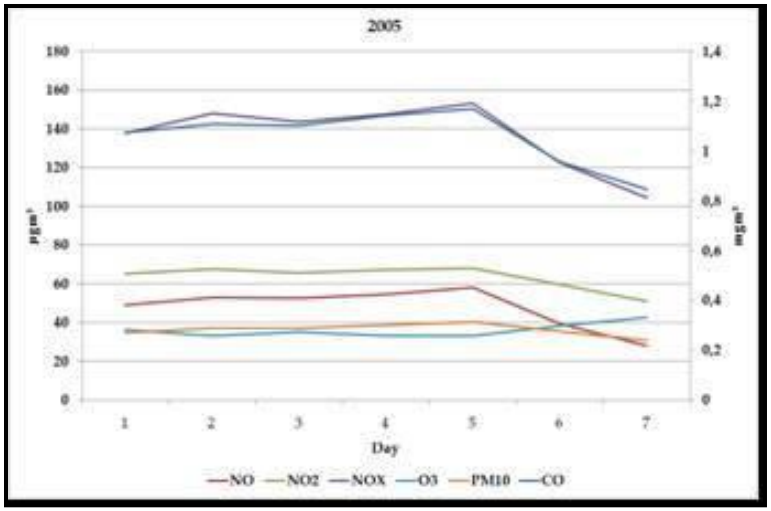


Time series

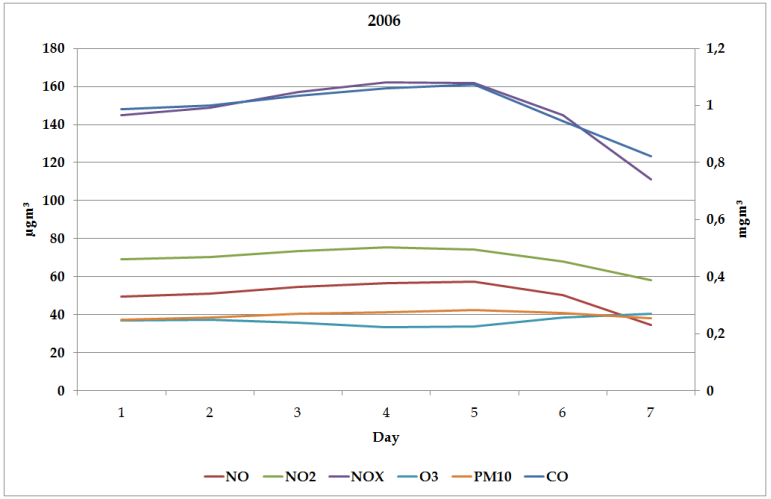
Day Type

ISTITUTO NAZIONALE PER L'ASSICURAZIONE CONTRO GLI INFORTUNI SUL LAVORO

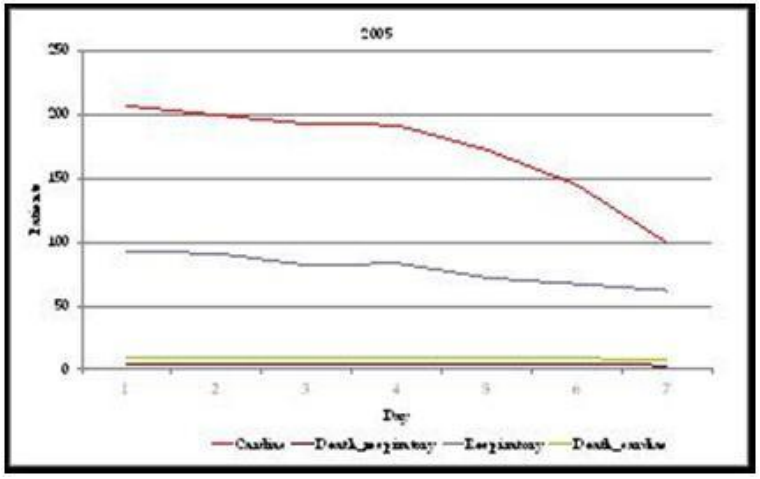
Weekly cycles for air pollution concentrations (2005)



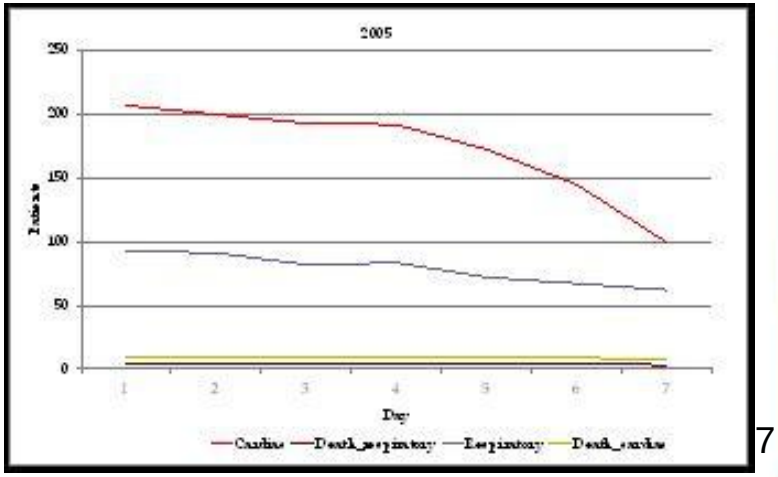
Weekly cycles for air pollution concentrations (2006)



Weekly cycles for cardio-respiratory admissions (2005)



Weekly cycles for cardio-respiratory admissions (2006)



METHODOLOGY

ISTITUTO NAZIONALE PER L'ASSICURAZIONE CONTRO GLI INFORTUNI SUL LAVORO

1-First- environmental time series

Ozone and Pm10 concentrations from meteorological conditions are a very important issues in air pollution

2- Pattern Selection

Patterns selection used for neural network is one of the most important tasks in order to achieve good generalisation

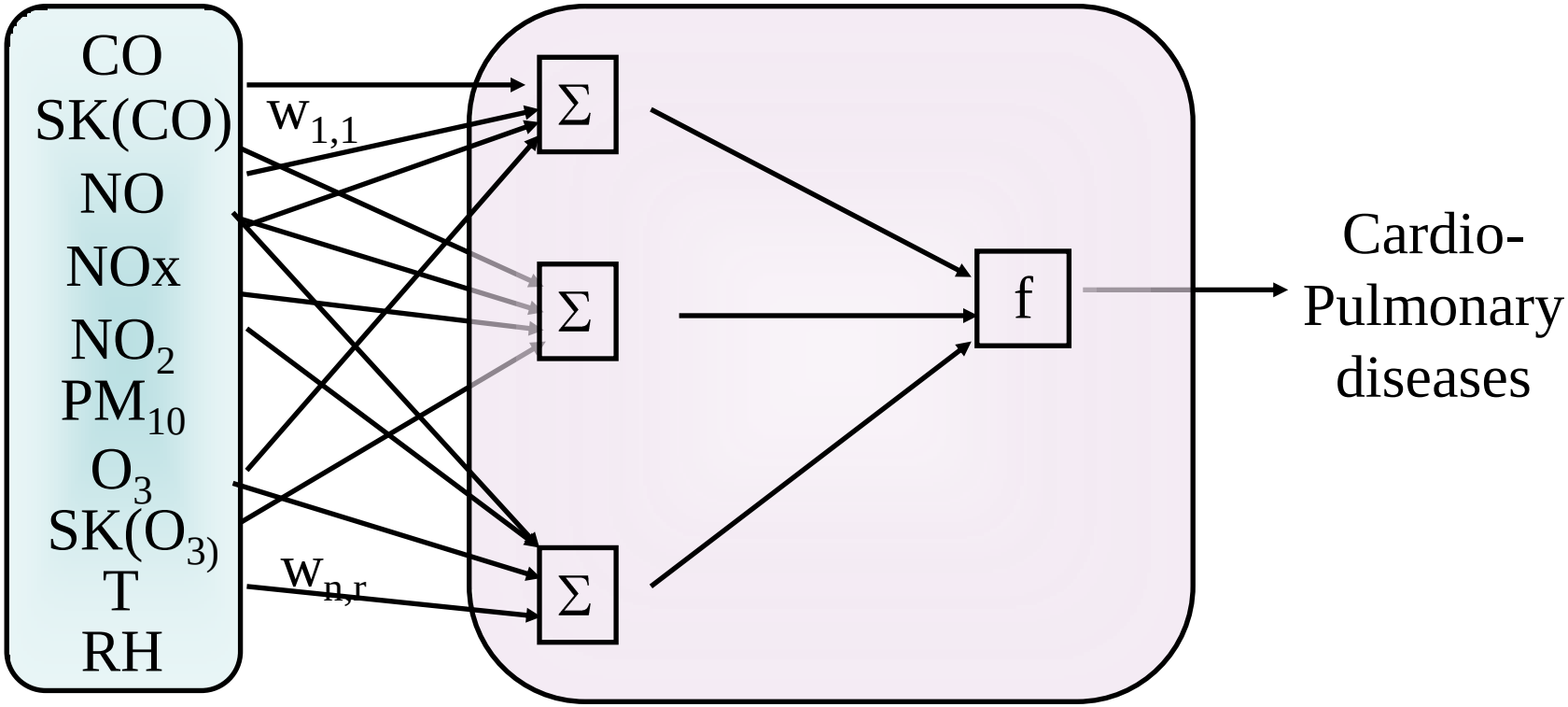
3-Neural Network

We apply NN to:

- examine the most injurious pollutants to human health and their mechanisms of dispersion
- identify factors and mechanisms of epidemic diffusion
- model the cardiovascular and respiratory diseases by means the use of air pollution data
- forecast five days in advance human health impact due to the different pollutants
- provide a reliable model for the prediction of the daily hospital admissions based on air quality and meteorological data.

Department of Production Plants and Human Settlements

NEURAL NETWORK ARCHITECTURE



Input set

Hidden layer: logistic

Output layer

Output

NEURAL NETWORK

Synthesis

ISTITUTO NAZIONALE PER L'ASSICURAZIONE CONTRO GLI INFORTUNI SUL LAVORO

Neural Network	
NEURAL NETWORK MODEL	MLP 6-10-1
HIDDEN NEURONS	6-9-12
ALGORITHM	CG-BFGS-GD
EPOCH	3000
ERROR FUNCTION	Sum of Square
HIDDEN ACTIVATION FUNCTION	LOGISTIC-TANH
OUTPUT ACTIVATION FUNCTION	IDENTITY
NETWORK RANDOMIZED	NORMAL

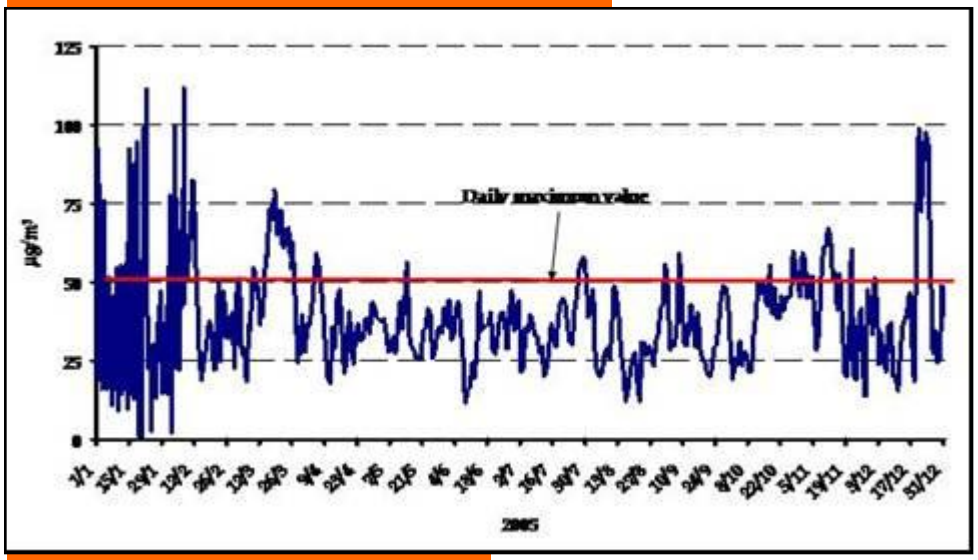
RESULTS AND DISCUSSION

Distribution

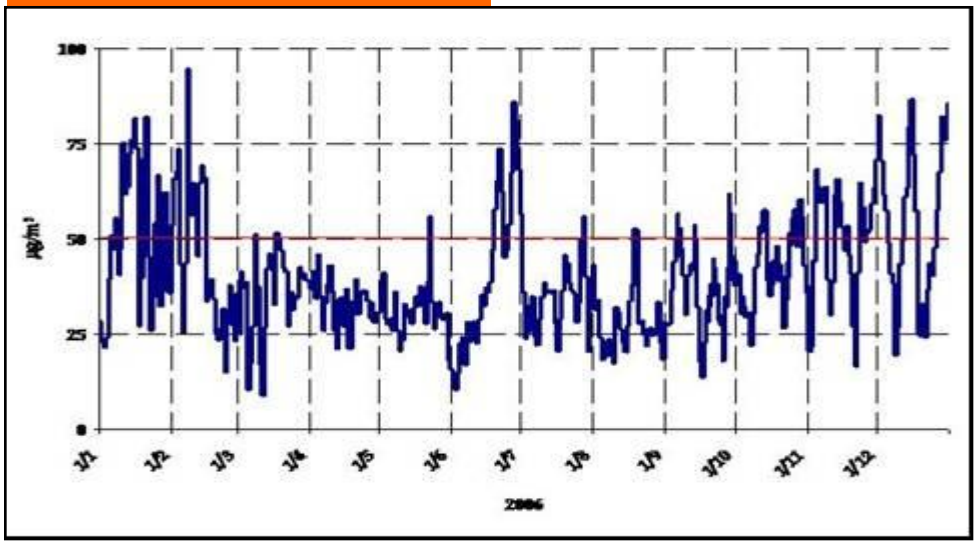
PM10 Distribution

INAIL
ISTITUTO NAZIONALE PER L'ASSICURAZIONE CONTRO GLI INFORTUNI SUL LAVORO

Pm10 trend (2005)



Pm10 trend (2006)

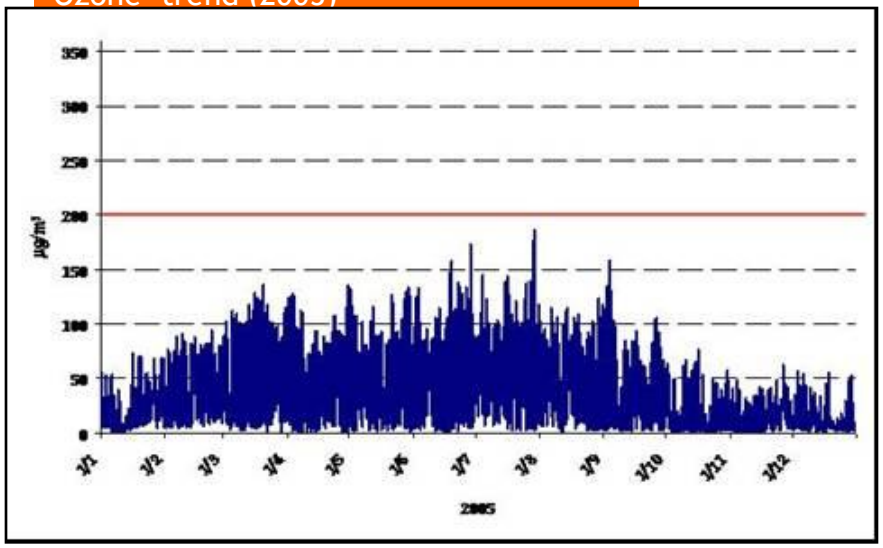


Distribution

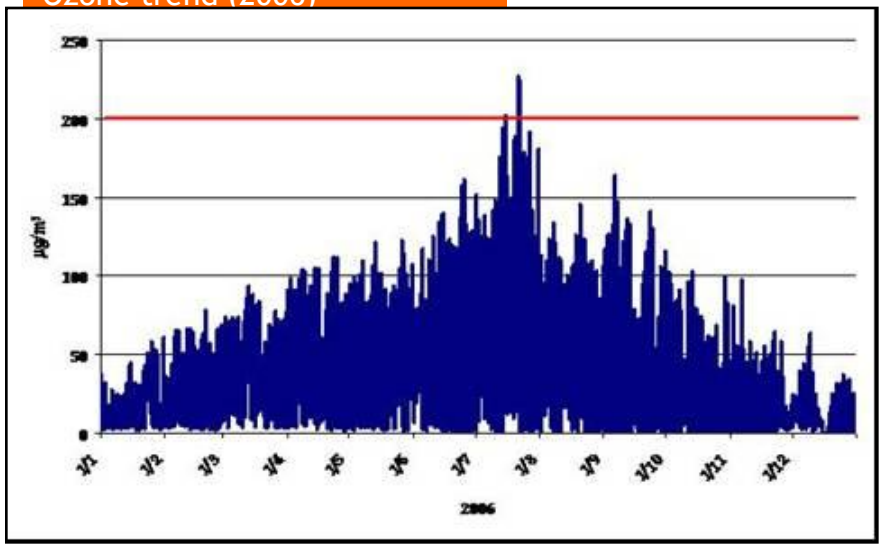
Ozone Distribution

ISTITUTO NAZIONALE PER L'ASSICURAZIONE CONTRO GLI INFORTUNI SUL LAVORO

Ozone trend (2005)



Ozone trend (2006)

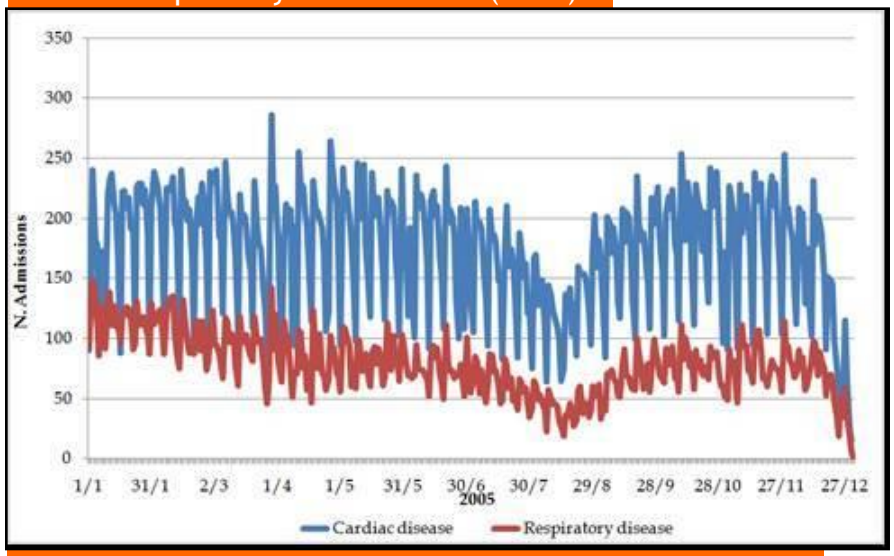


Distribution

Cardio-respiratory disease trend

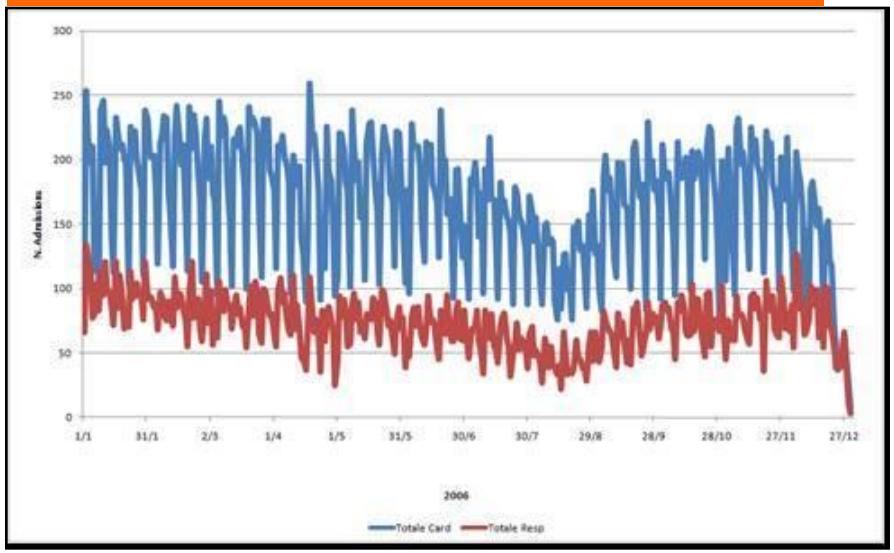
ISTITUTO NAZIONALE PER L'ASSICURAZIONE CONTRO GLI INFORTUNI SUL LAVORO

Cardio-respiratory disease trend(2005)



- The admissions for cardiac disease rang from 100-250 patients
- Patients with respiratory disease rang from 0-150 patients

Cardio-respiratory disease trend trend (2006)



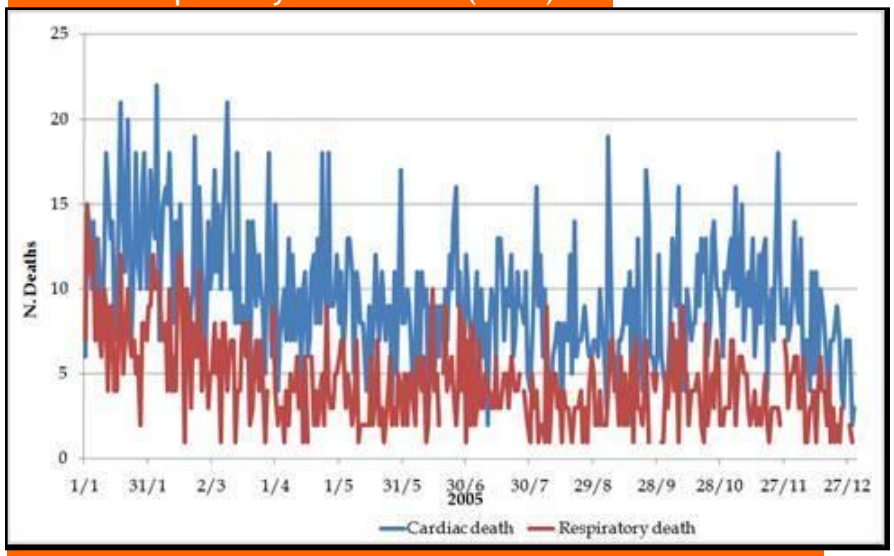
- peak at the end of April

Distribution

Cardio-respiratory death trend

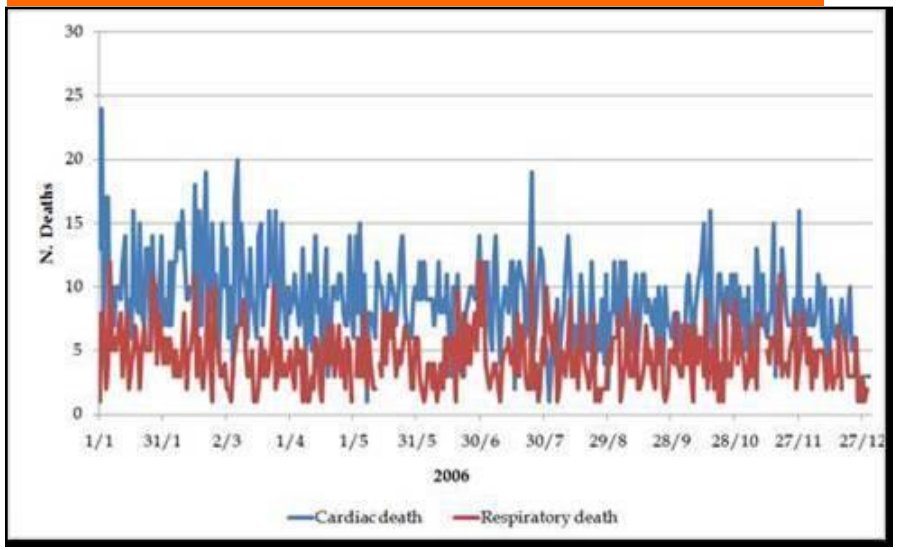
INAIL
ISTITUTO NAZIONALE PER L'ASSICURAZIONE CONTRO GLI INFORTUNI SUL LAVORO

Cardio-respiratory death trend(2005)



- The deaths from respiratory disease do not exceed 15 units

Cardio-respiratory death trend trend (2006)



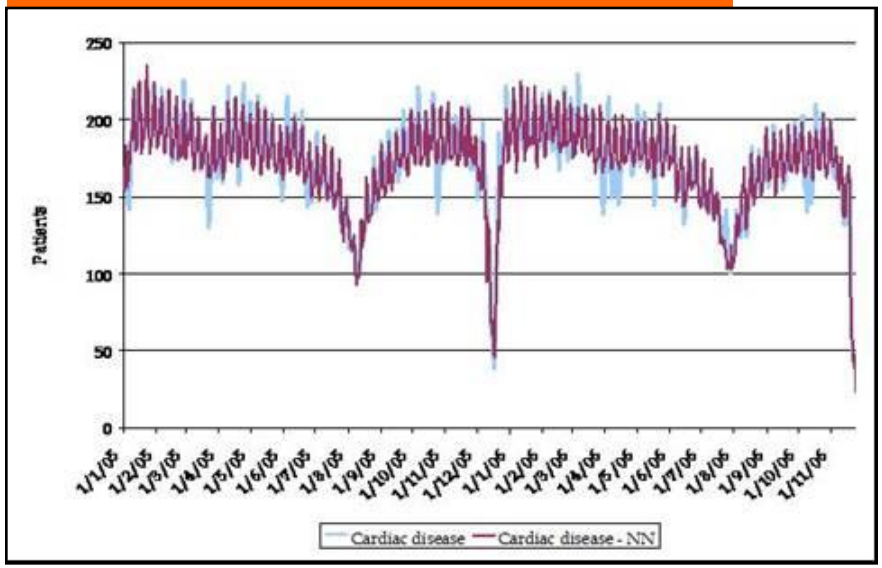
- The deaths from respiratory disease do not exceed 15 units

NN RESULTS

Cardio-respiratory admissions

ISTITUTO NAZIONALE PER L'ASSICURAZIONE CONTRO GLI INFORTUNI SUL LAVORO

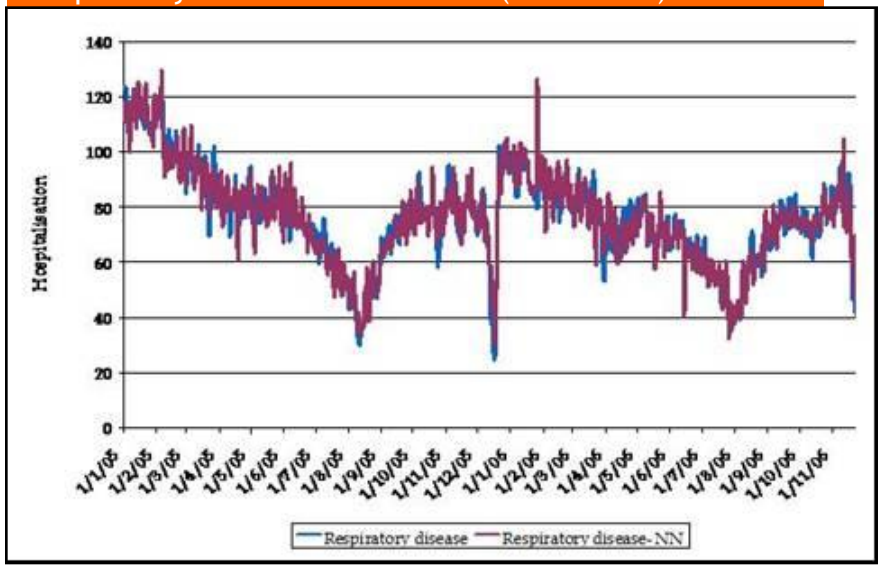
Cardiac admissions for Rome (2005-2006)



- The determination coefficient is very high (R2=0.93)

Cardiac hospitalisation by NN	
Percent of wrong simulation	9.5
Percent of right simulation	90.5

Respiratory admissions for Rome (2005-2006)

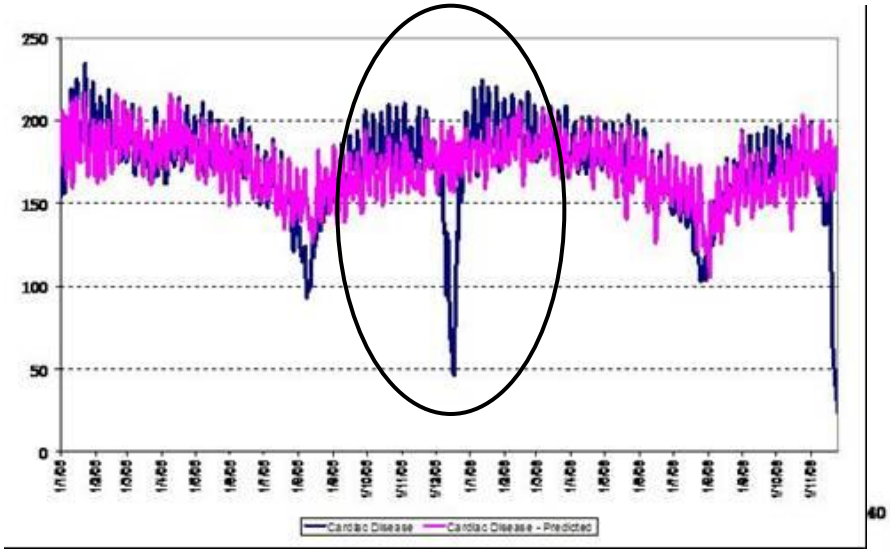


- The determination coefficient is a little worse respect to previous simulation (R2=0.92), even if levels are very significant

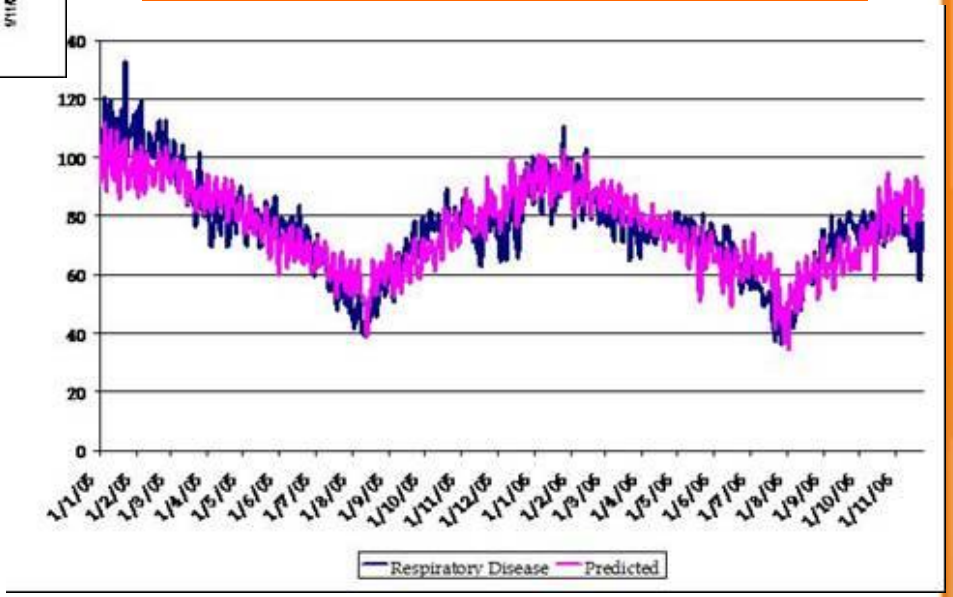
Respiratory hospitalisation by NN	
Percent of wrong simulation	17.9
Percent of right simulation	82.1

Regression Stastical Model

Cardiac admissions for Rome (2005-2006)



Respiratory admissions for Rome (2005-2006)



REGRESSIVE RESULTS

Cardio-respiratory admissions

INAIL
ISTITUTO NAZIONALE PER L'ASSICURAZIONE CONTRO GLI INFORTUNI SUL LAVORO

Training phase		
	Cardiac Admissions	**Respiratory Admissions**
R2	**0,44**	**0,78**
Correlation coefficient	0.67	0.88
Mean absolute error	0.25	0.27
Root mean squared error	0.43	0.32
Relative absolute error	62.32 %	53.28%
Root relative squared error	74.60 %	46.79 %
Total Number of patterns	451	451
Test phase		
	Cardiac Admissions	**Respiratory Admissions**
R2	**0,31**	**0,765**
Correlation coefficient	0.557	0.875
Mean absolute error	0.271	0.261
Root mean squared error	0.449	0.317
Relative absolute error	73.14 %	54.13%
Root relative squared error	83.34%	48.43%
Total Number of patterns	242	242

Conclusions

INAIL
ISTITUTO NAZIONALE PER L'ASSICURAZIONE CONTRO GLI INFORTUNI SUL LAVORO

❶ These first results show that NN applied to time series is able to model the relations between the environment and meteorological data and health effects on the populations

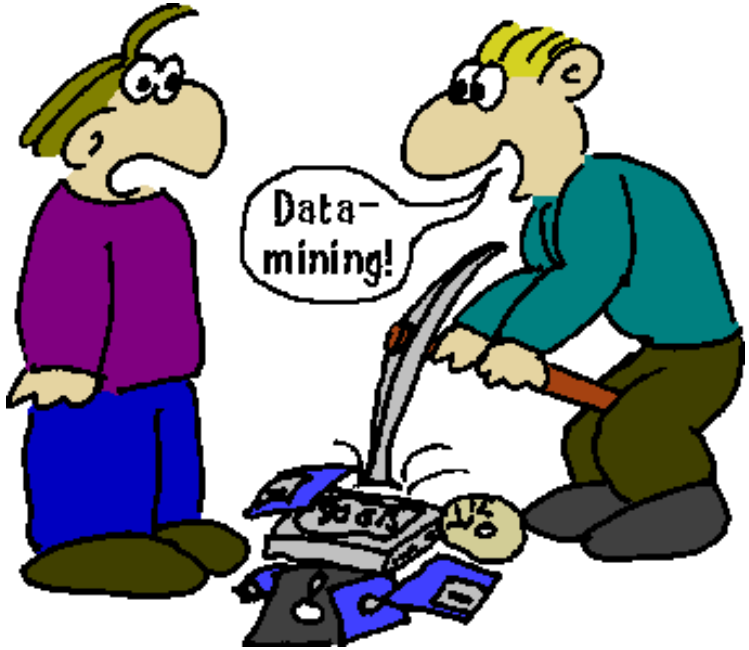
❷ The pre-processing analysis shows that the **skewness** coefficients for the pollutants can be giving a more accurate connection with the real human exposure

❸ These first results showed the importance of the environmental-epidemiological problem and the major areas in order to forecast with some accuracy the short-term effects on human health of the environmental component

❸ By using NN, it was possible to determine the association statistically meaningful between the cases of death or hospitalization and pollutants (Pm10 and/or O3)

INAIL
ISTITUTO NAZIONALE PER L'ASSICURAZIONE CONTRO GLI INFORTUNI SUL LAVORO

Thanks for your attention



Questions or Comments?

Email: Armando.pelliccioni@ispesl.it