Slide 1

Forecasting High Ozone Levels Using Networks of Active Neurons

Rodrigo Cerda H.
Héctor Jorquera G.

Santiago, September 29th 1999

Slide 2

General Aspects

Slide 3

Atmospheric Pollution

Direct effect on people's life quality

London December 1952: more than 4.000 people died

Main cause of atmospheric pollution

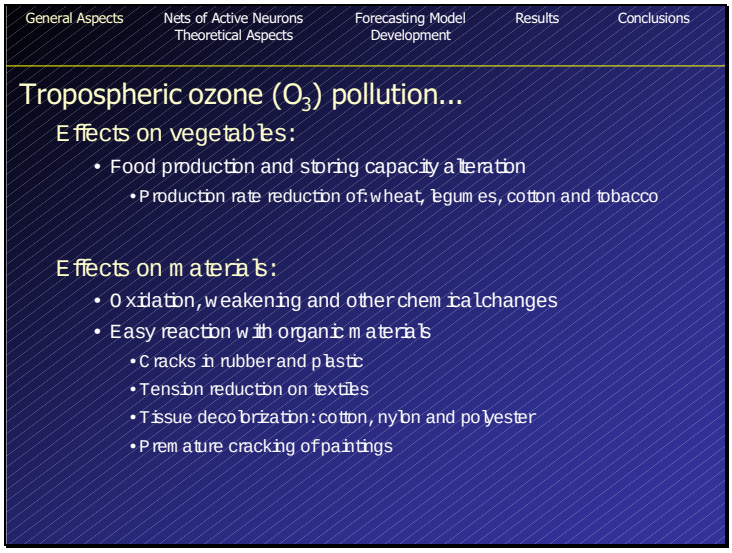
Massive usage of fossil fuels

Tropospheric ozone (O_3) pollution

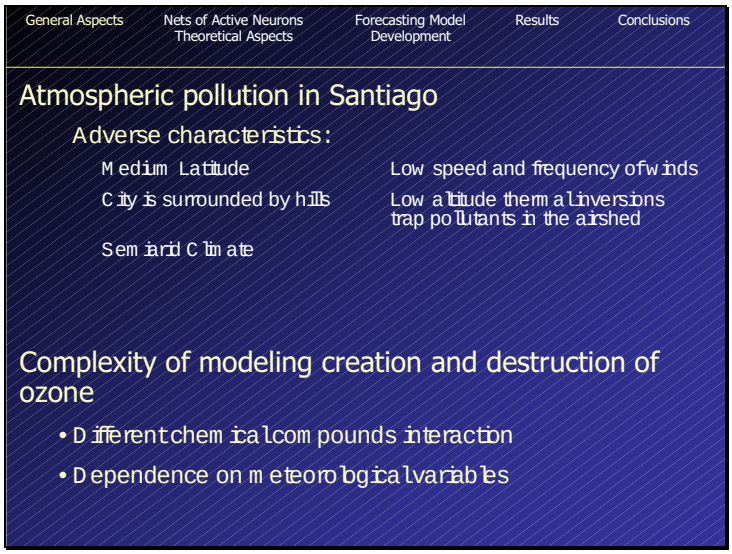
Effects on population:

- Breathing system mucous membrane irritation
  - Cough, nausea, breathing difficulties, lung tissue irritation and diminishment of breathing capacity

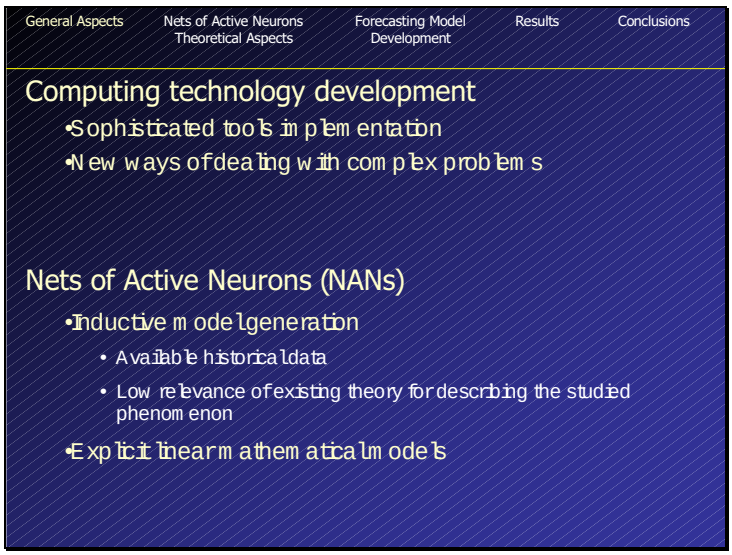
Slide 4



Slide 5



Slide 6



Slide 7

Research objective

Analyze the ability of a Net of Active Neurons of developing a maximum ozone concentration forecasting model for Santiago city

Slide 8

Nets of Active Neuron's Theoretical Background

Slide 9

Complex object modeling

Deductive methods:

- Based on object theory and physical principles
- Widely used in sciences and engineering
- Limited to relatively simple problems or objects with well known theory

Inductive methods:

- Based on available historical data
- Do not require knowledge of the object's theory
- Best for modeling complex objects or problems for which there is no identified theory

Slide 10

Group Method of Data Handling (GMDH) algorithms (active neurons)

- Generate inductive models
- Invented in 1967 by A.G. Ivakhnenko (Ukraine)

Basic characteristics of GMDH algorithms

- Based on Adaptive Networks
  - Identify the optimal structure during the parameter estimation process
- Carry out an ordered search of the optimal complexity model which define the studied object

Slide 11

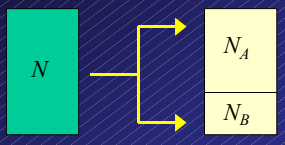
GMDH Algorithms Classification

	Parametric	Non-parametric
Continuous variables	Combinatorial (COMBI) Multilayered Iterational (MIA) Objetive System Analysis (OSA) Harmonical Two-level (ARIMAD) Multiplicative-Additive	Objective Computer Clusterization (OCC) Pointing Finger Clusterization Algorithm (PF) Analogues Complexing (AC)
Binary variables	Harmonical Rediscretization	Multilayerd Theory of Statistical Decisions (MTSD)

Slide 12

Combinatorial Algorithm (COMBI)

- Parametric algorithm (coefficient estimation)
- Based on the Least Squares Method:
  - Objective variable must be identified explicitly
  - Only continuos objective variables
- Best models are selected based on an external criterion :

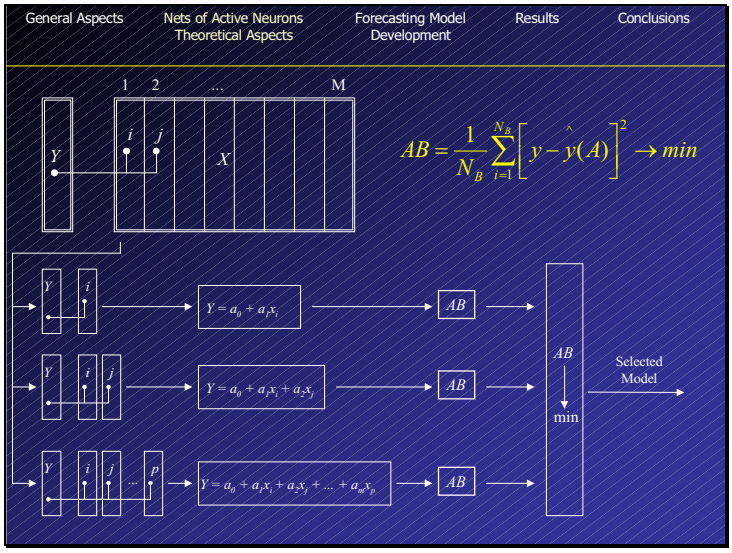


Learning Sub-sample (coefficient estimation)

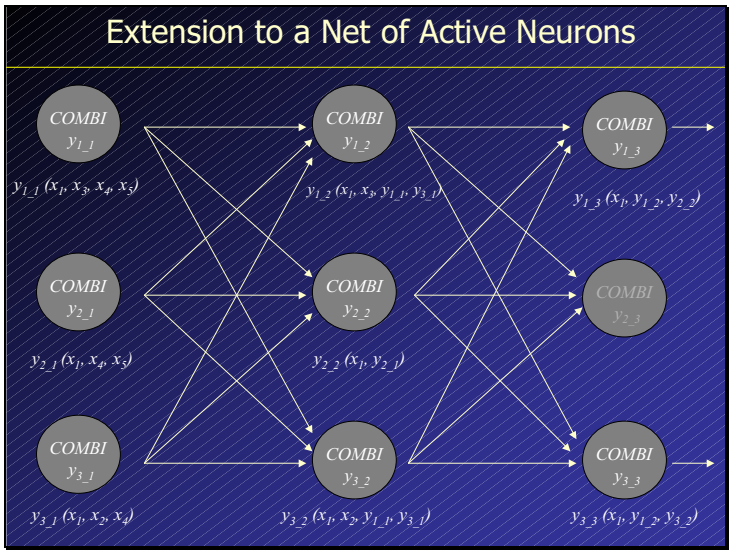
Training Sub-sample (model evaluation)

Dispersion with respect to Y

Slide 13



Slide 14



Slide 15

Slide 16

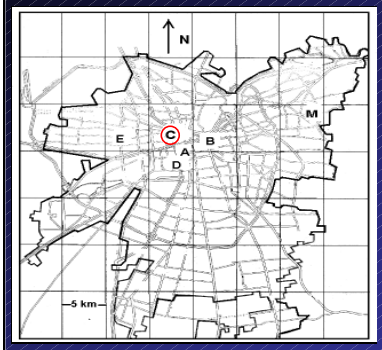
Software implementation

- Programs developed in MATLAB 5.2 programming language
- GMDH algorithms developed:
  - Combinatorial (COMBI)
  - Multilayered Iterative Algorithm (MIA)
  - Analogues Complexing (AC)

Slide 17

Historical data used

Automatic Air Quality Monitoring Network (MACAM)



Meteorological Variables :

Temperature: maximum, minimum, standard deviation

Wind speed and direction

Air relative humidity

...

Pollutants:

SO_2 HCT

CO NO_x

O_3

Slide 18

Basic Set Generation → **Processing** → **Input Variables Selection**

- Station C historical data
- 2558 daily observations in 23 original variables (1988 -1994)
- Irrelevant variables suppression (7)
- Missing information filtering (Microsoft Access)

→ 506 observations in 16 variables

Slide 19

Basic Variables

Symbol	Unit	Variable
CO_MX_t	[ppm]	Maximum matinal carbon monoxide concentration
DR_PR_t	Sexagesimal degrees	Daily average of wind direction
HR_PR_t	[%]	Daily average of air relative humidity
HR_STD_t	[%]	Standar deviation of relative air humidity
$NO2\_MX_t$	[ppb]	Maximum matinal nitrogen dioxide concentration
NO_MX_t	[ppb]	Maximum matinal nitric oxide concentration
NOX_MX_t	[ppb]	Maximum matinal NOx concentration
$O3\_MX_t$	[ppb]	Daily maximum ozone concentration
SP_PR_t	[m/s]	Daily average of wind direction
$THETA_t$	Sexagesimal degrees	Daily dispersion of wind direction
T_MN_t	[°C]	Minimum daily temperature
T_MX_t	[°C]	Maximum daily temperature
$TO3\_MX_t$	[°C]	Temperature at moment of highest ozone concentration
T_PR_t	[°C]	Average of daily temperature
T_STD_t	[°C]	Standar deviation of daily average temperature
V_PR_t	[m/s]	Daily average of wind direction

Slide 20

Basic Set Generation → **Processing** → **Input Variables Selection**

- Include time lags (t, t-1, t-2 y t-3)
- Outliers elimination:

$$|x_i - \mu_x| \geq 3\sigma_x$$

313 observations for 16x4 = 64 lagged variables

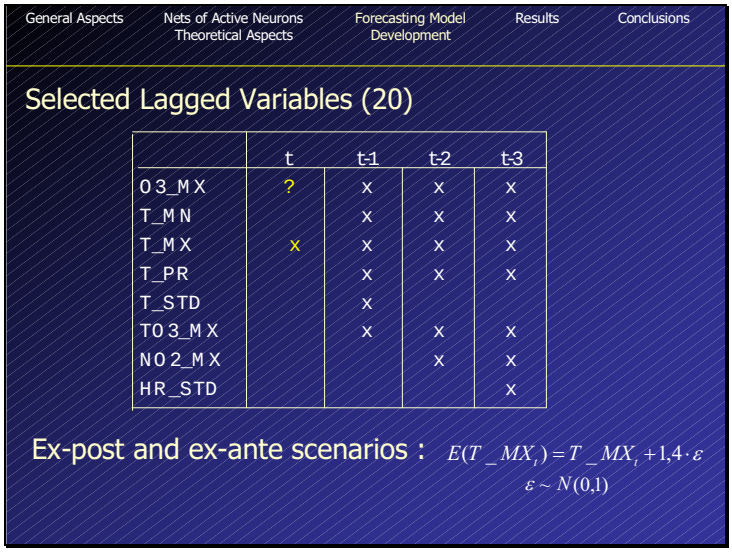
Slide 21

Basic Set Generation → **Processing** → **Input Variables Selection**

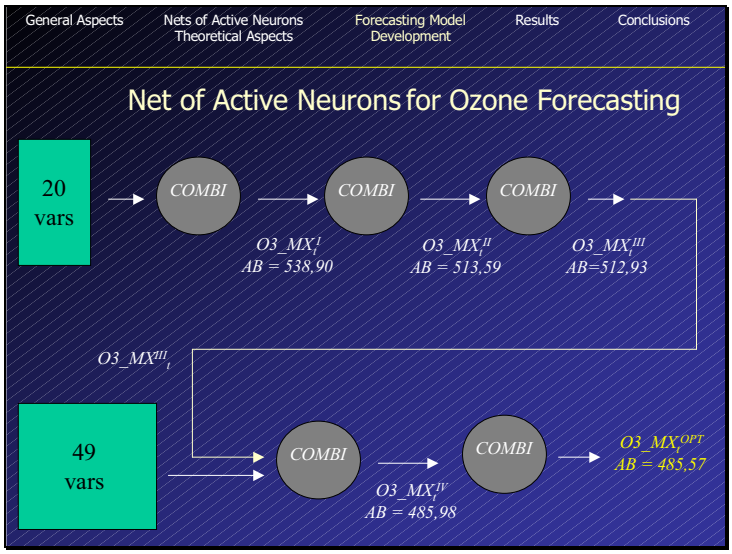
- Limited computing capacity
- Heuristics methods used:
  - Correlation analysis based selection
  - Partial results quality

Combinatorial Algorithm
20 input variables

Slide 22



Slide 23



Slide 24

Slide 25

Optimal model for ozone forecasting

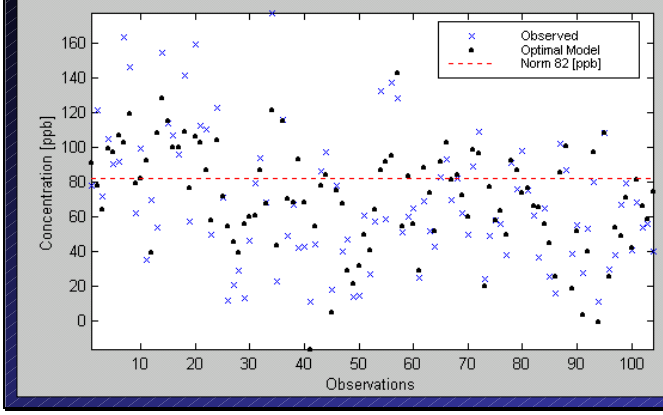
	T_MX	*T_MN*	*T_PR*	*T_STD*	*TO3_MX*	*HR_PR*
t	7,2394					
t-1	-4,4727	-1,0652	2,4047	-3,4444		
t-2	-2,2414	-1,6923	0,9423		-0,0284	
t-3		0,1198	0,2458			0,2572

	SP_PR	*DR_PR*	*O3_MX*	*NO_MX*	*NO2_MX*	*CO_MX*
t						
t-1	0,3611	-0,0108	0,3154	0,0121	0,0723	
t-2		-0,0323	0,2353	0,0210		
t-3		0,0668	0,1286		-0,0769	-0,6103

$a_0 = -34{,}7869$

Slide 26

Optimal model in the ex-post scenario



Slide 27

Alternative deductive models

- **Persistence:**
  - Incapacity of identify the causes of the studied phenomenon
  - Actual concentration level is a good estimate

$$O3\_MX_t = O3\_MX_{t-1}$$

- **Linear Deductive:**
  - Analysis of the laws and principles involved in ozone generation and destruction
  - Atmospheric Diffusion Equation

$$O3\_MX^t = \alpha + \beta \cdot O3\_MX^{t-1} + \gamma \cdot T\_MX^{t-1} + \delta \cdot T\_MX$$

$$O3\_MX^t = 9.5936 + 0{,}5309 \cdot O3\_MX^{t-1} - 5{,}4743 \cdot T\_MX^{t-1} + 6{,}4645 \cdot T\_MX^t$$

Slide 28

General Aspects | Nets of Active Neurons Theoretical Aspects | Forecasting Model Development | Results | Conclusions

Information used in results evaluation

External data set

Was not used in parameters estimation

Validity of the obtained results

Slide 29

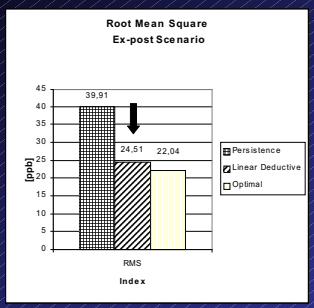
Forecasting Capacity

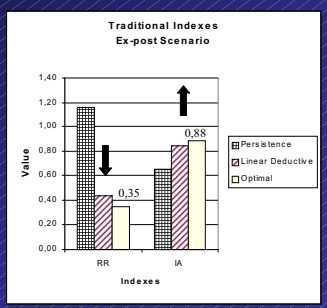
Slide 30

General Aspects | Nets of Active Neurons Theoretical Aspects | Forecasting Model Development | Results | Conclusions

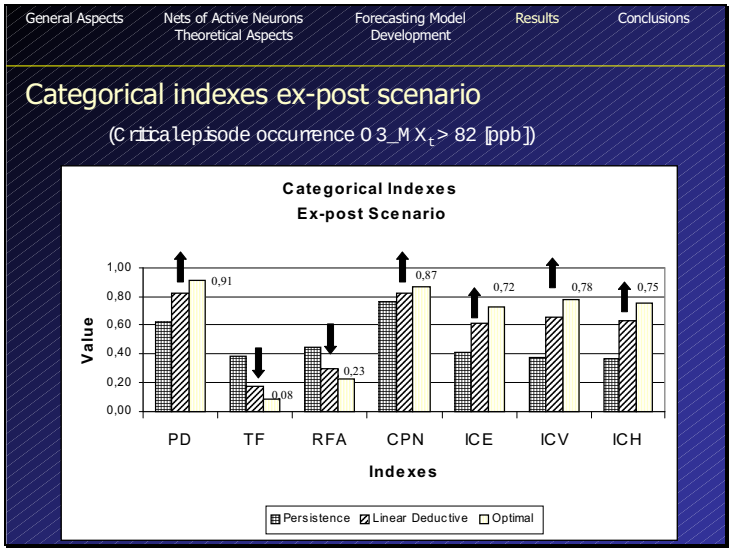
Traditional indexes ex-post scenario

(Observed and forecasted values similarity)

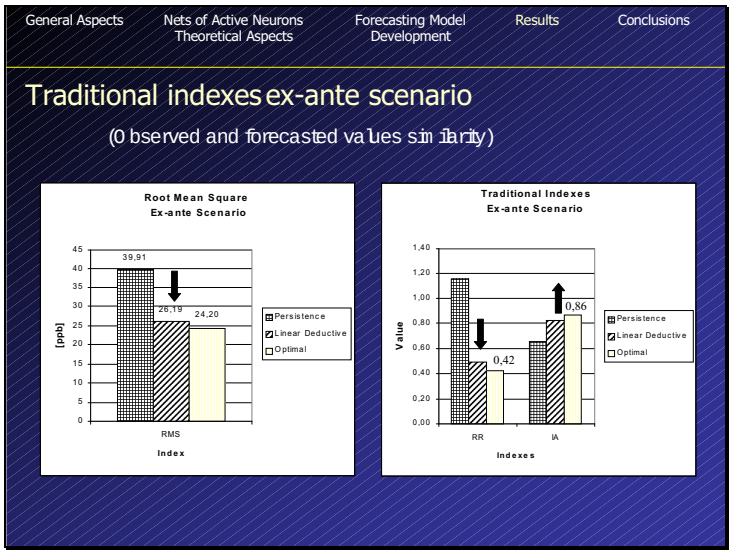




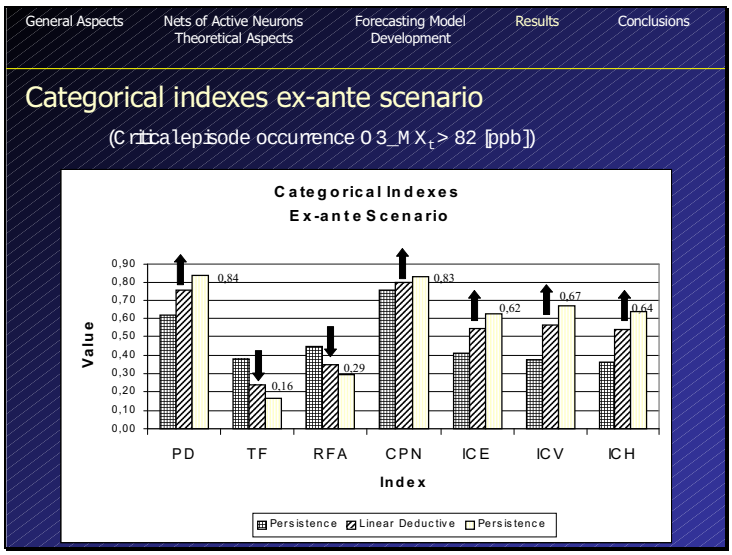
Slide 31



Slide 32

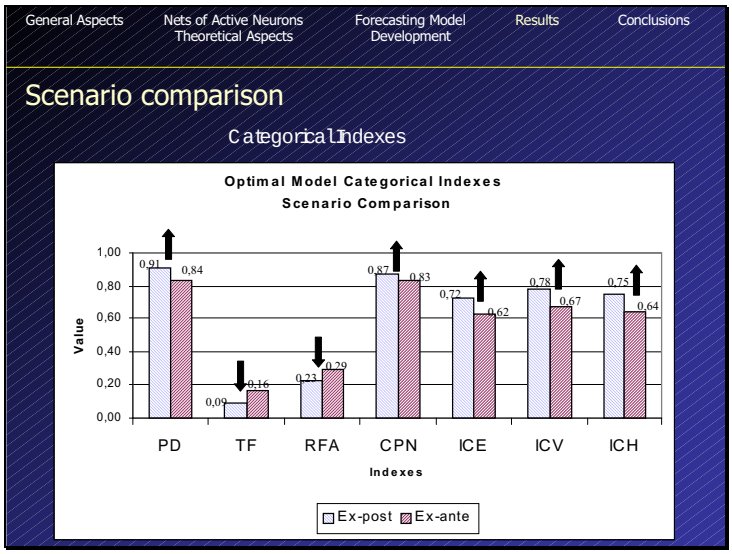


Slide 33

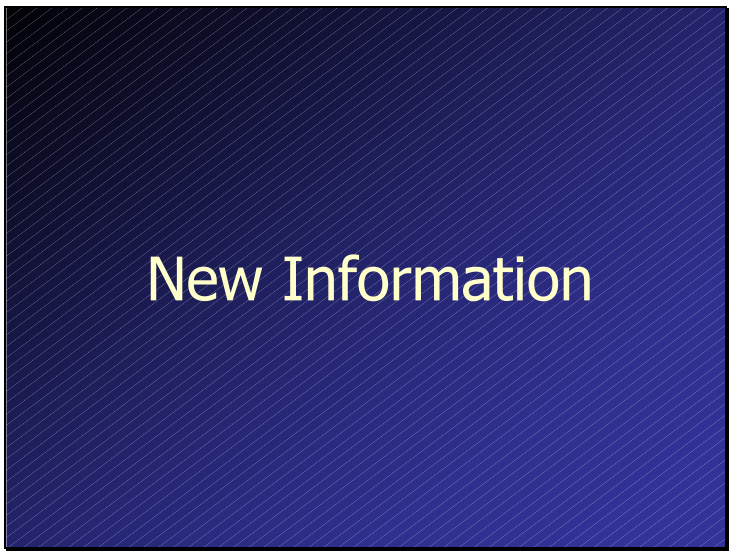


Slide 34

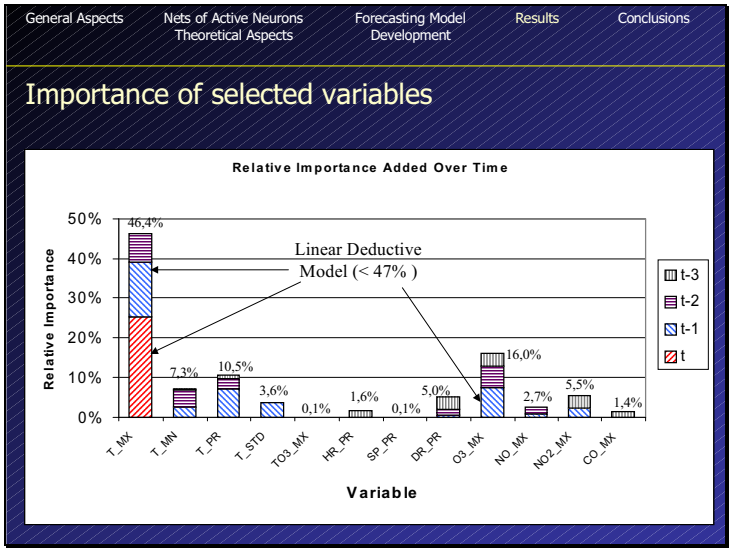
Slide 35



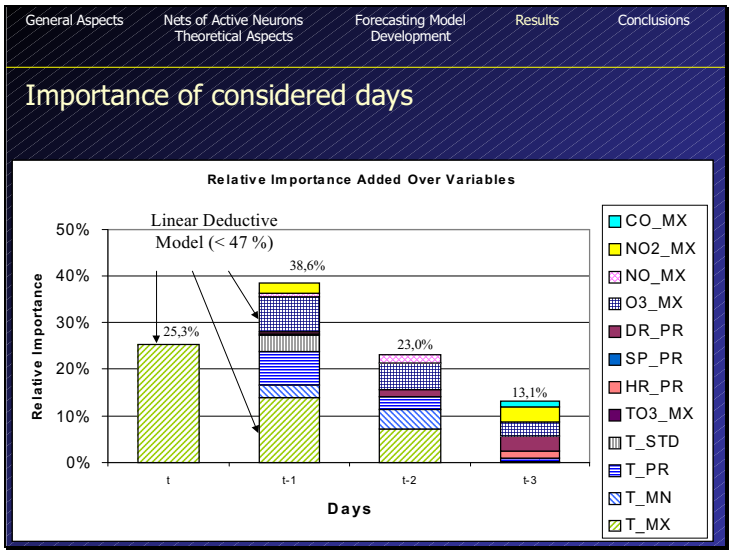
Slide 36



Slide 37



Slide 38



Slide 39

Slide 40

Nets of Active Neurons

- Adequate tool for modeling and forecasting O_3
- Forecasting precision is very good
- Explicit model generated is an important additional result

Forecasting models development using NAN

- Selection of an appropriate algorithm
- Selection of an appropriate input variable set

Slide 41

Further research

Ozone pollution

- Usage of additional inductive modeling tools for improving forecasting precision (Neural Networks)
- Deductive analysis of the explicit model generated

Nets of Active Neurons

- Usage of alternative algorithms (MIA, AC or others)
- Improve the software design and implementation

Slide 42

Forecasting High Ozone Levels Using Networks of Active Neurons

Rodrigo Cerda H.
Héctor Jorquera G.

Santiago, September 29th 1999