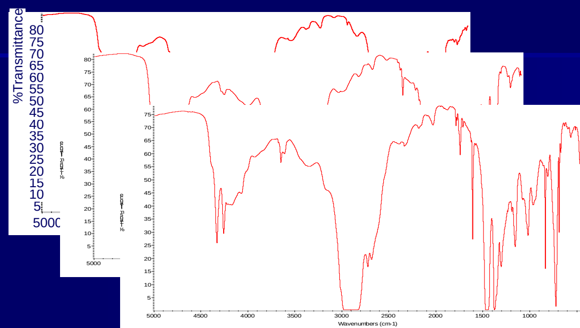
Propylene Concentration Prediction in Ethylene-Propylene Co-polymers by Partial Least Squares, Principal Component Regression and Artificial Neural Networks

Emilio Marengo, Marco Bobba, Elisa Robotti

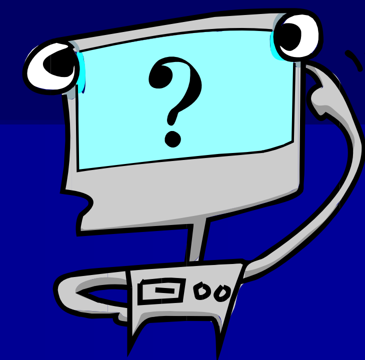
University of Eastern Piedmont
Department of Environmental and Life Sciences
Via Bellini 25/G - 15100 Alessandria

Purpose

88 ethylene – propylene co-polymers



Propylene Concentration

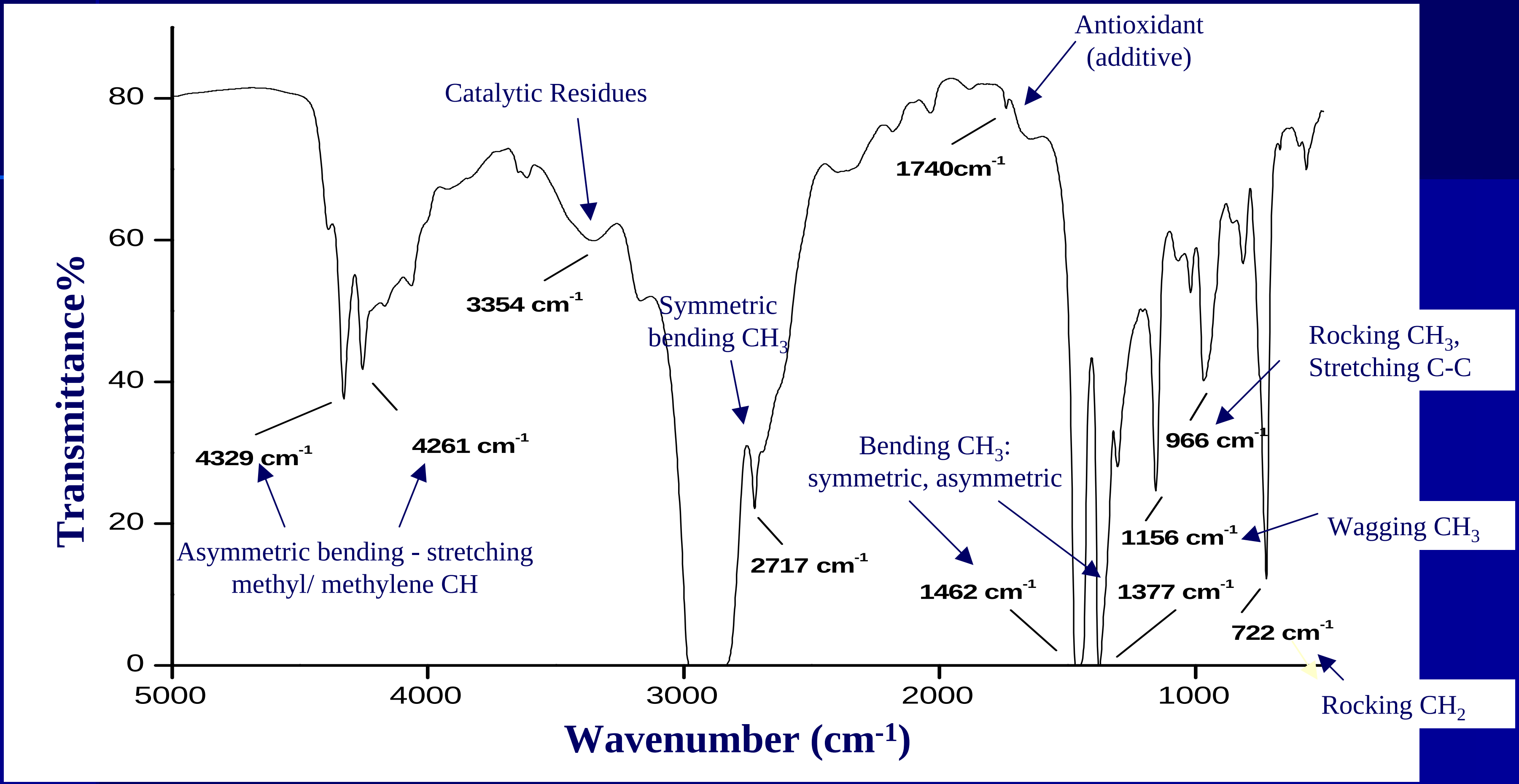


Partial Least Squares Regression (PLS)

Principal Component Regression (PCR)

Artificial Neural Networks (ANNs)

Dataset



Resolution: 4 cm^{-1}
Scans: 16
Range: 5000 - 450 cm^{-1}
Detectos: DTGS
Sample: Pressed/melted pellet

1° DERIVATIVE

Validation Procedure

PCR PLS ANNs → Validation procedure?

- PCR, PLS ↓ Cross validation
- ANNs ↓ Training / Test Set splitting

88 samples {

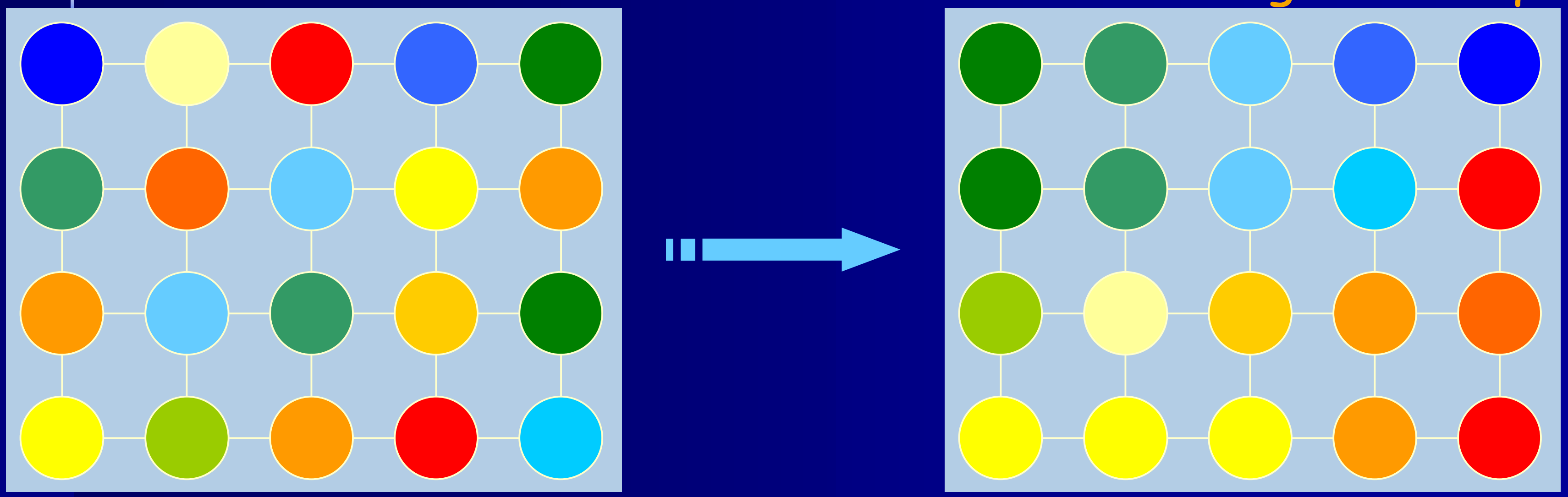
- **Training set** To build the model
- **Test Set** To end the training phase
- **Production Set** For genuine prediction

KOHONEN ARTIFICIAL NEURAL NETWORK

Kohonen SOMs (Self Organising Maps)

Unsupervised artificial neural networks can group the objects of a dataset into different classes, on the basis of their similarity

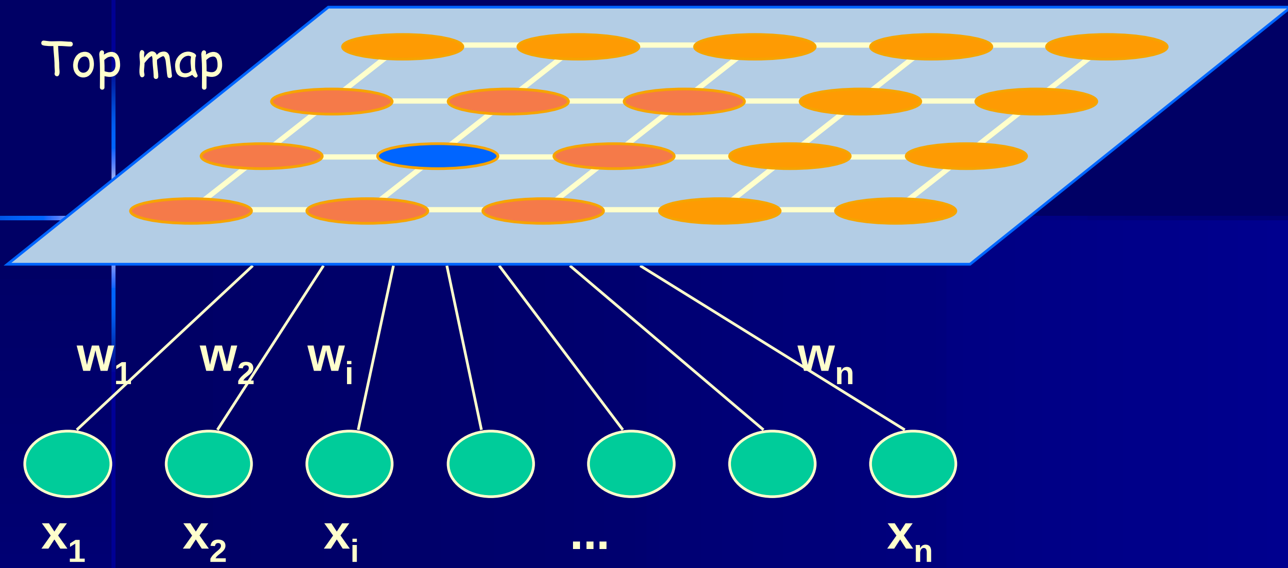
Used here instead of experimental design techniques



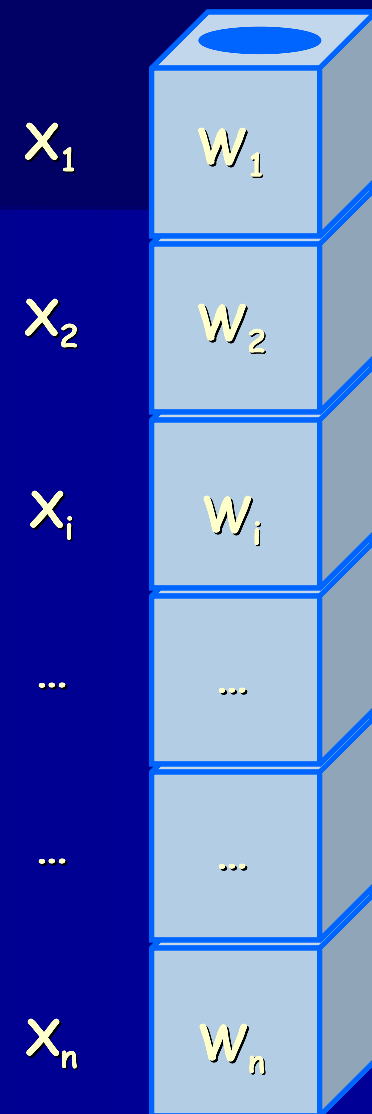
Similar objects are contained in the same neuron or in adjacent ones on the top map

The weights on the layers below the top map indicate the reasons for their similarity

Kohonen SOMs (Self Organising Maps)



Input neurons



$$\text{Winning Neuron} \leftarrow \min\left\{\sum_{i=1}^{n}(x_i - w_i)^2\right\}$$

$$w_i = w_i^{prec} + \eta \times a\left(d_{vinc} - d_j\right) \times \left(x_i - w_i^{prec}\right)$$

Learning rate

Function of the distance (here: linear)

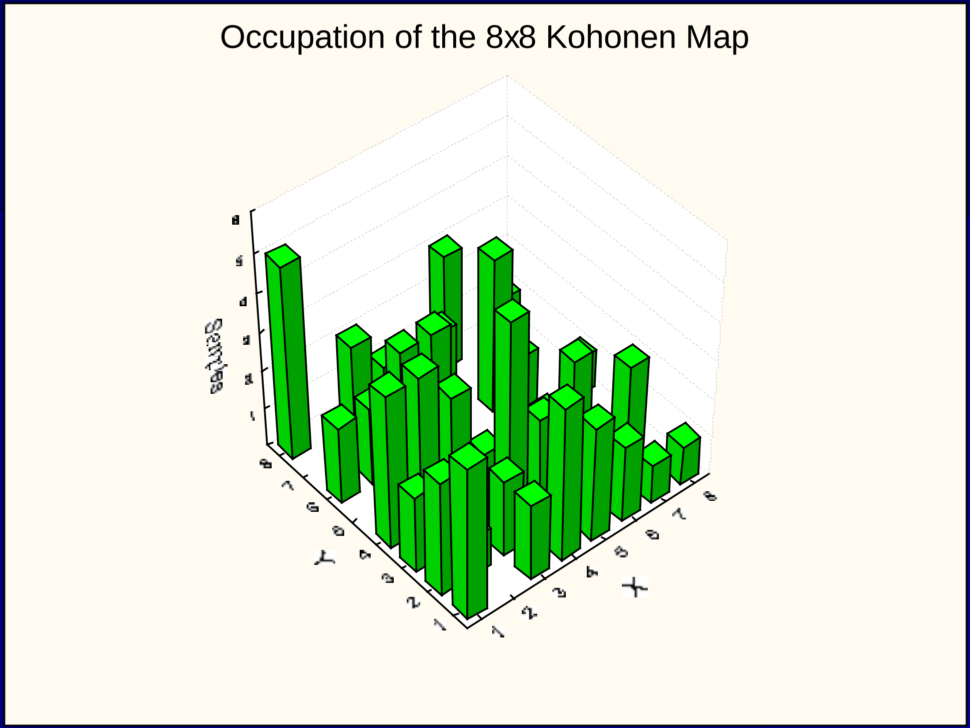
Distance from the winning neuron

Kohonen Maps

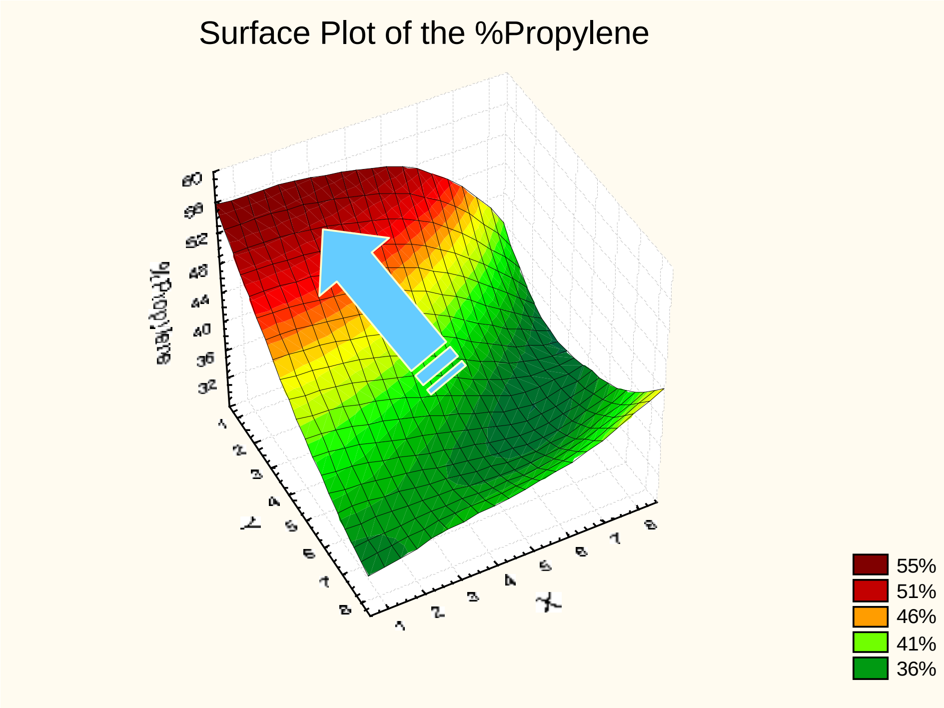
Cell occupied by 1 sample
⇓
training set

Cell occupied by >1 sample
⇓
2 samples with extreme values of exp. response (plus eventually an intermediate value) for the training set

Samples not selected for Tr are selected for test or production sets



Top layer: 8 x 8 neurons
Iterations: 100
Distance: Euclidean
Threshold Error: 0.001
Correction propagation: 8 – 0
Learning Rate: 0.8 – 0.01 (linear)



Kohonen network built only on the spectral data

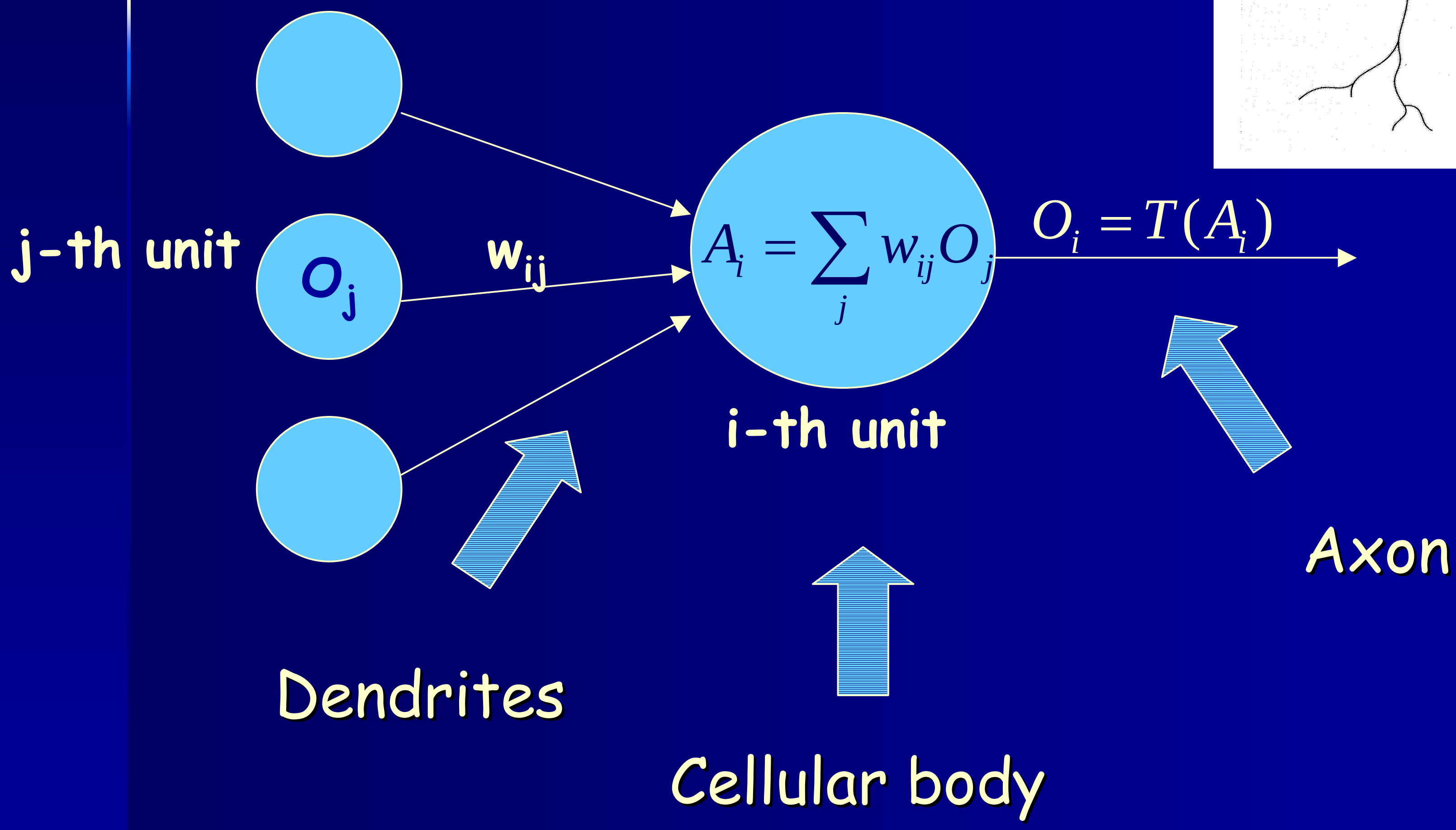
Clustering of the samples according to the propylene concentration

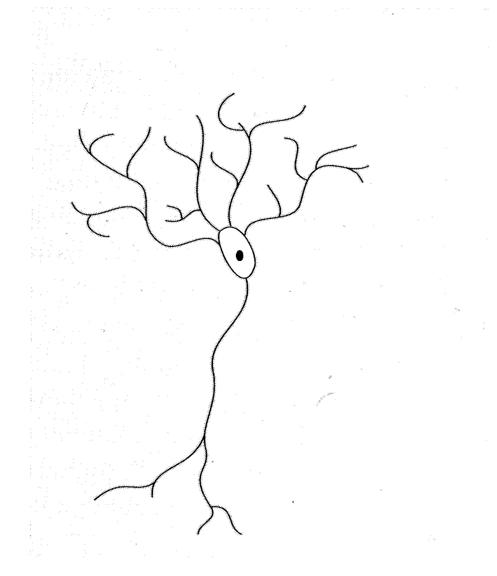
The selection of the training set from different cells of the top map guarantees a homogeneous representation of the samples

Initial dataset	Training set	Test Set	Production set
88 samples	**55** samples	**20** samples	**13** samples

Artificial Neural Networks

From human to artificial neurons…

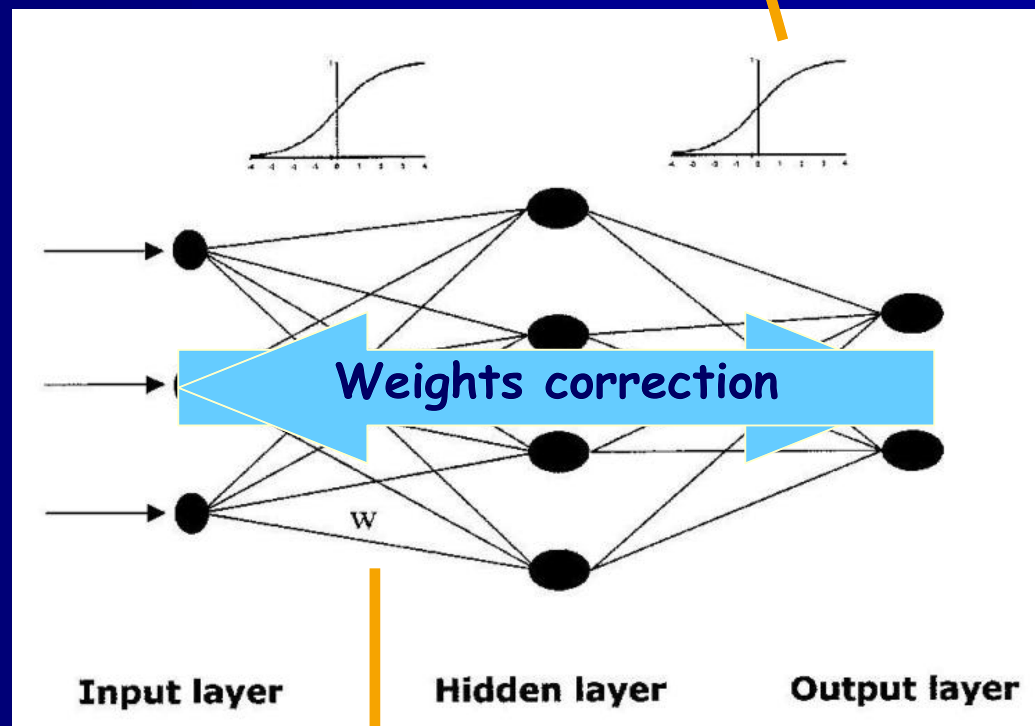




Feed-Forward Back-Propagation Artificial Neural Networks

Each layer connected to the following through a selected transfer function

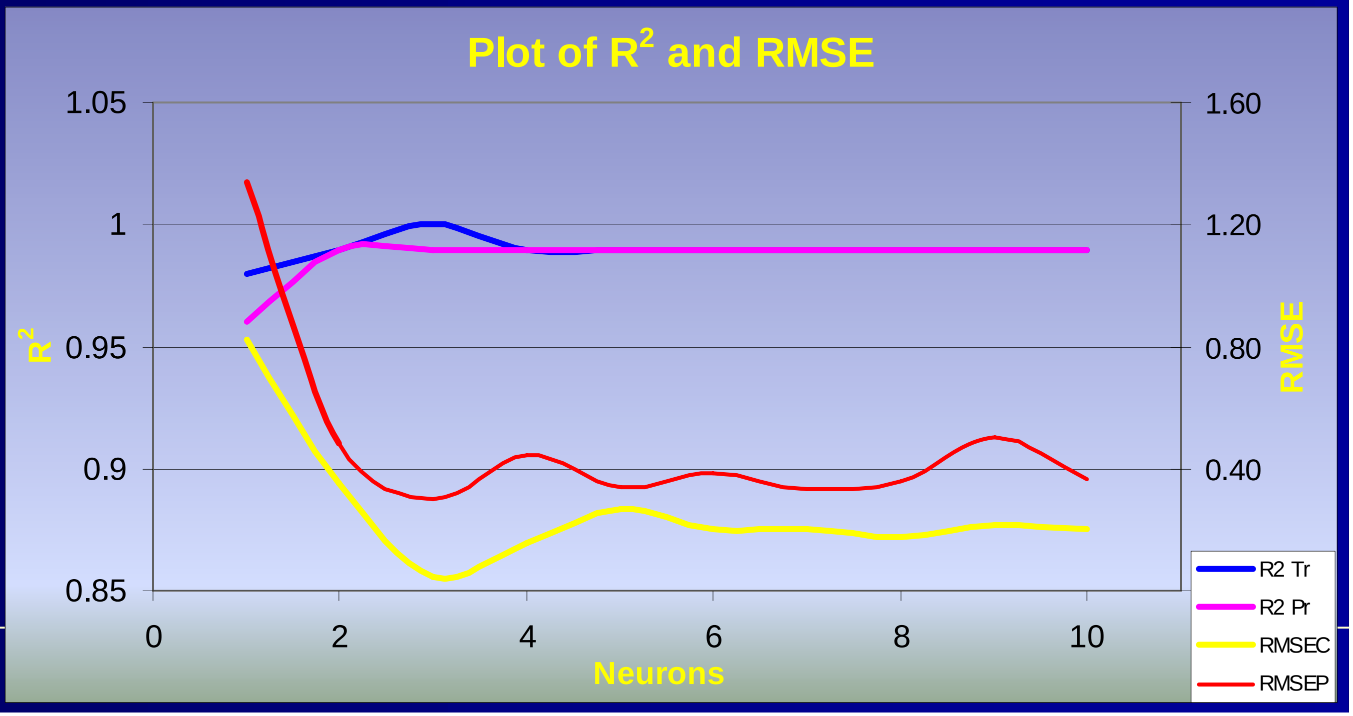
- **Input layer**
  Each neuron associated to an experimental variable
- **Hidden layer(s)**
  Variable number of neurons
- **Output Layer**
  Each neuron associated to a response

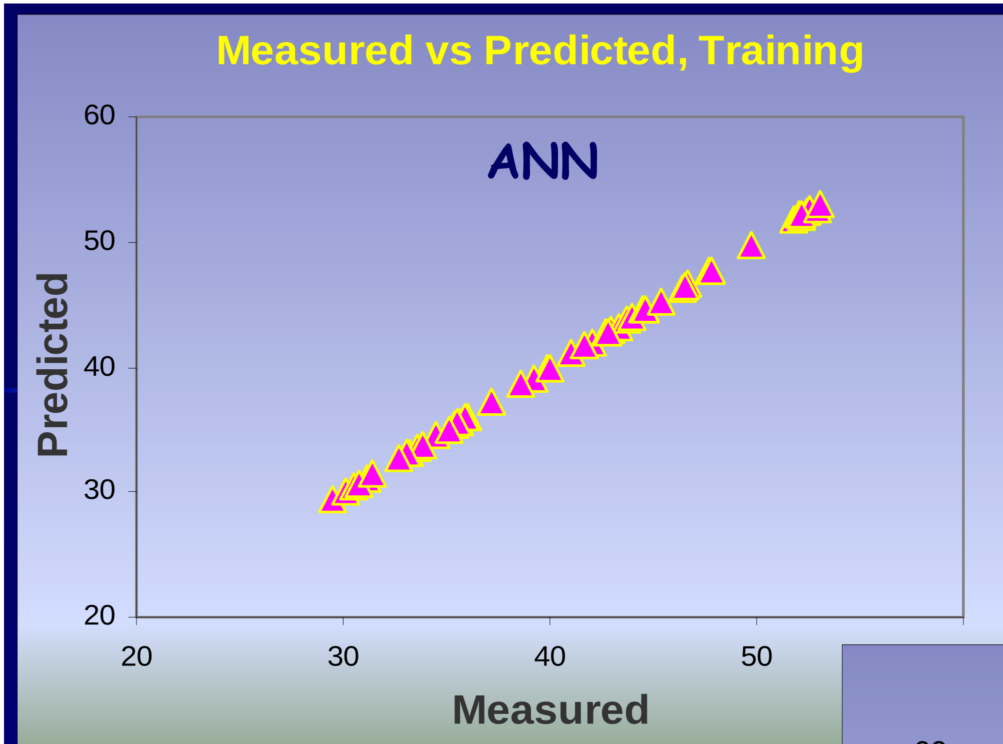


Neurons connected through weights

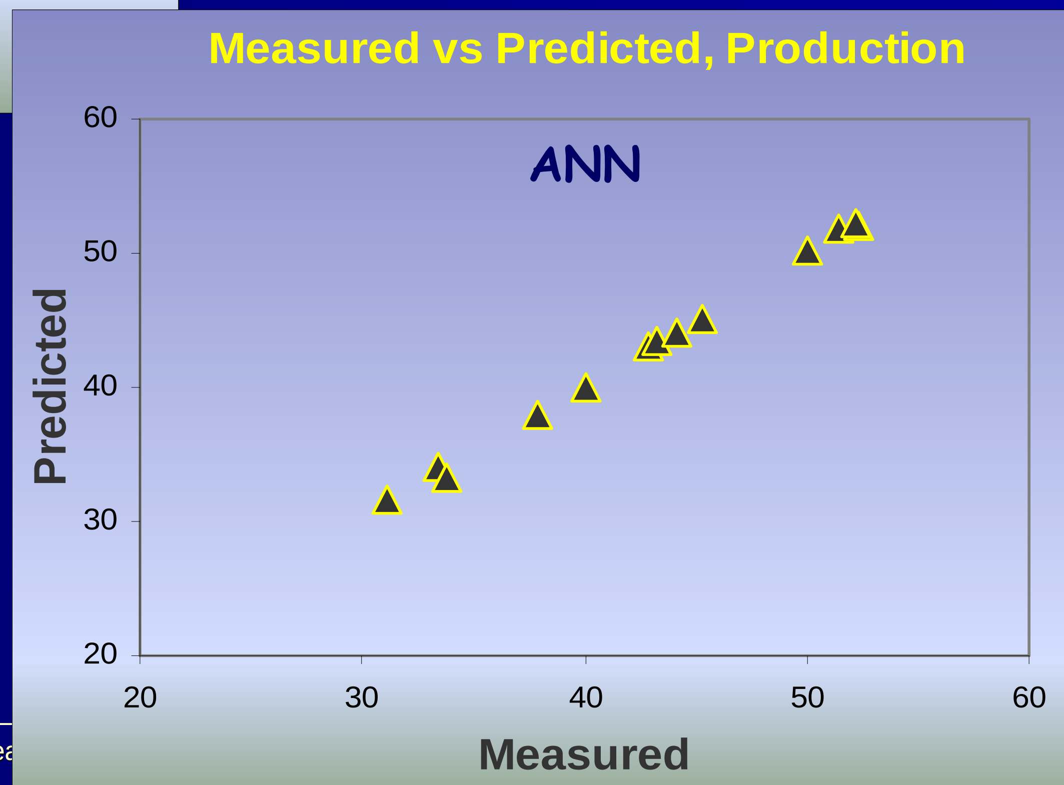
Results: Feed-Forward Back-Propagation Artificial Neural Networks

Input layer	Hidden layer	Output layer	Momentum / Learning rate
236 Neurons Range Scaling [-1; 1]	1 layer with 3 neurons logistic transfer function	1 neuron logistic transfer function	0.3 / 0.3





	R^2	RMSE
Training Set	0.99	0.04
Test Set	0.99	0.30
Production Set	0.99	0.30

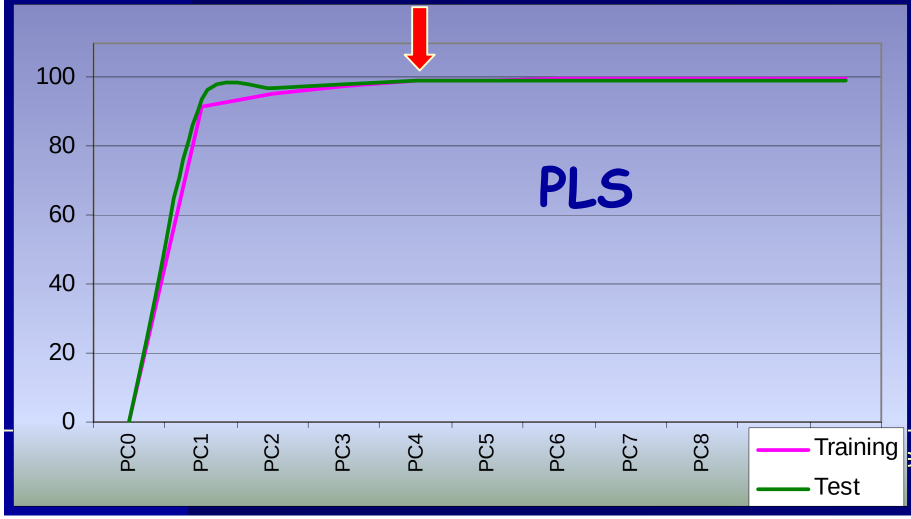


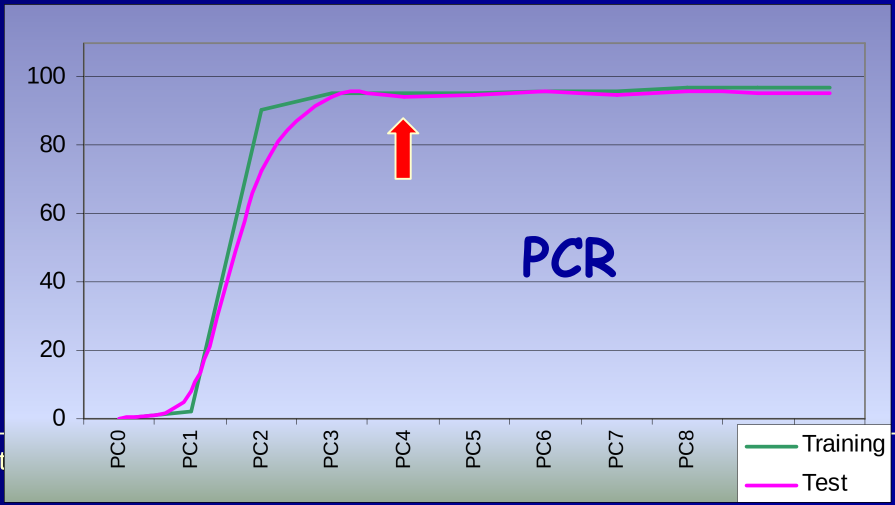
Optimal number of PCs ➡ determined by the Test set
Predictive ability of the model ➡ tested on the Production set

Results
PLS - PCR

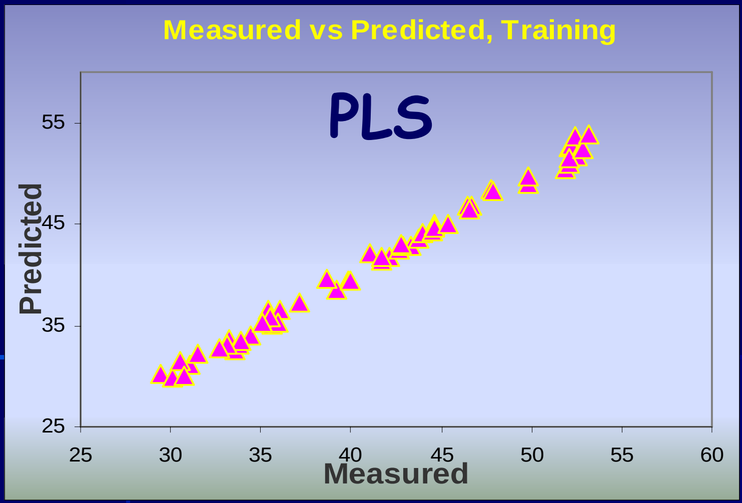
	% Expl. Var. Tr set	% Expl. Var. Ts set
PC1	91.67	83.46
PC2	95.59	90.32
PC3	97.27	97.45
PC4	98.93	98.27
PC5	99.30	98.74
PC6	99.57	98.95
PC7	99.69	99.51
PC8	99.78	99.44
PC9	99.85	99.44
PC10	99.90	99.55

	% Expl. Var. Tr set	% Expl. Var. Ts set
PC1	2.38	8.36
PC2	90.30	72.38
PC3	95.40	94.07
PC4	95.52	94.51
PC5	95.54	94.66
PC6	95.65	96.05
PC7	95.71	94.82
PC8	97.02	95.70
PC9	97.04	95.22
PC10	97.14	95.55

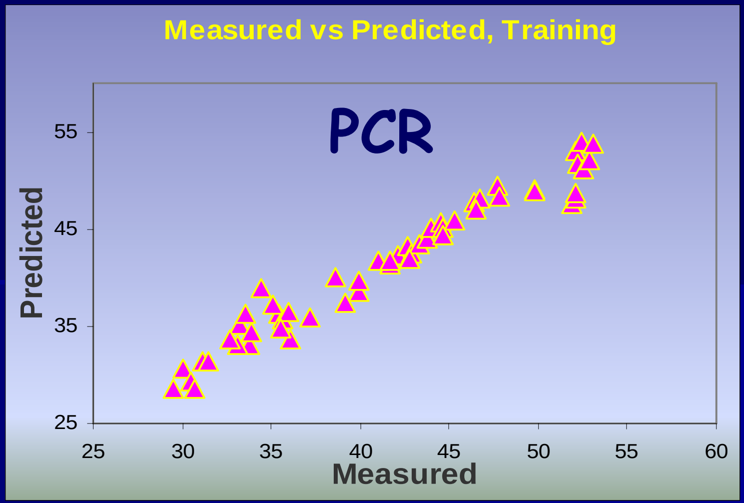




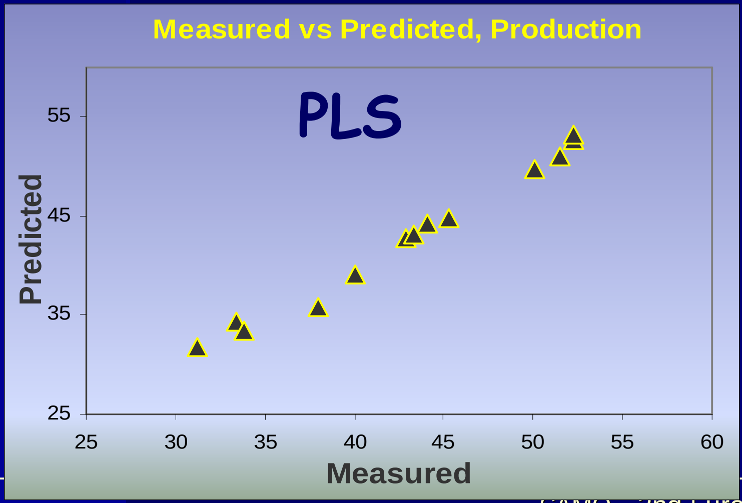
Results: PLS - PCR

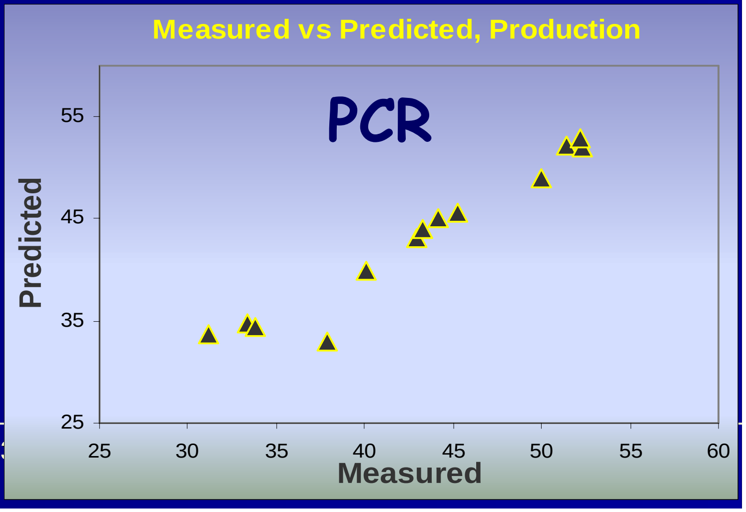


	R^2	RMSE
Training Set	0.99	0.60
Test Set	0.99	0.66
Production Set	0.98	0.81



	R2	RMSE
Training set	0.95	1.52
Test set	0.96	1.38
Production set	0.94	1.69





COMPARISON

	Fitting Ability		Predictive Ability	
	R^2 Tr	RMSEC	R^2 Pr	RMSEP
ANN	0.99	0.04	0.99	0.30
PLS 4PCs	0.99	0.60	0.98	0.81
PCR 4PCs	0.95	1.52	0.94	1.69

CONCLUSIONS

- ✓ **Artificial Neural Network** is the method that provides the **best fitting** and, most of all, **the best predictive ability**, if compared with PCR and PLS . The results are very satisfactory so the ANN model can be successfully used to predict the future propylene concentrations in ethylene-propylene co-polymers.
- ✓ **PLS** and **PCR** algorithms show performances, both in fitting and prediction, worse than those obtained with the artificial neural network, **but PLS works better respect PCR**.
- ✓ **Kohonen network** proved to be an **efficient procedure for selecting the training set**, in fact it mapped the spectra of samples having similar propylene concentration to the same neuron.

Acknowledgements

Polimeri Europa S.P.A. - Ferrara (Italy)

- Dr. Silvana Di Martino

Chemometrics Group
Dept. Environmental and Life Sciences
Alessandria (Italy)

-Prof. Emilio Marengo
-Dr. Marco Bobba
-Dott. Valentina Longo
-Dott. Marco DeMartini