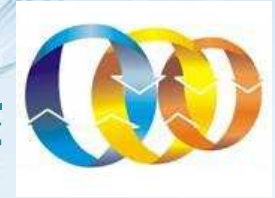




**UNIVERSIDAD AUTÓNOMA CHAPINGO**



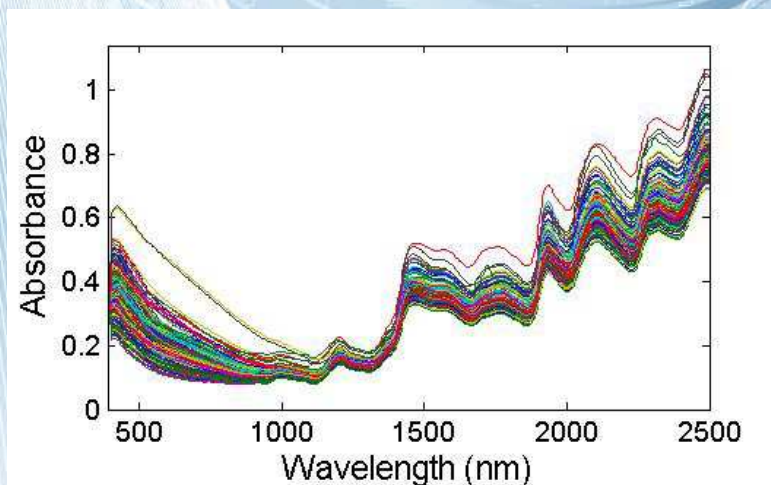
**12th European Symposium on Statistical Methods for the Food Industry**

# **Regression methods with $p$ large to predict chemical components**

**Carmen Ybarra-Moncada**  
Paris, France. The 29<sup>th</sup> of February 2012

## Motivation

- How do I know the quality attributes of my potato harvest?
- Prediction of future  $y$  values given future  $x$  values
- Chemical analysis are time consuming and expensive
- NIRS data for  $n$  potatoes
- $p$  large
- The  $x$ -variables may be highly correlated

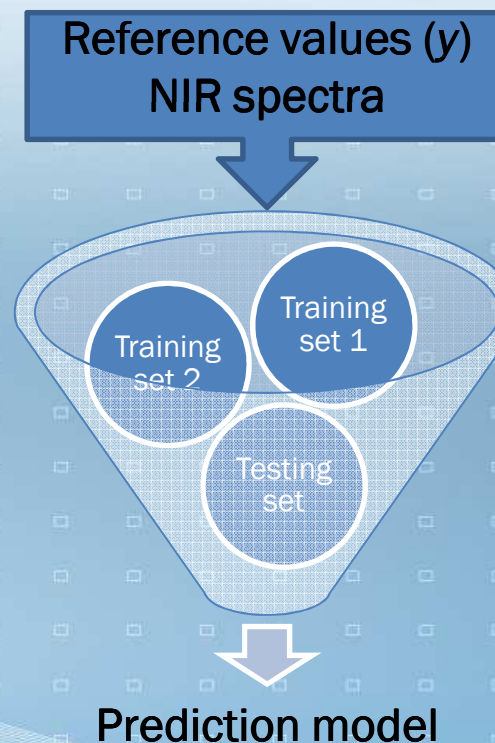


NIR spectra of 220 flour samples made from potato clones (400-2500 nm).

## Methodology

Regression methods

- PLSR
- PCR
- LS-SVM
- LS-SVM<sub>PC</sub>
- LS-SVM<sub>PLS</sub>





## Results

- PCR and PLSR outperform the LS-SVM method
- The LS-SVM<sub>PC</sub> and LS-SVM<sub>PLS</sub> approach improved the performance in prediction of LS-SVM
- Frequently LS-SVM predicted any y-value as the mean of the y-values in the training set
- Eleven potato clones with large starch content

Method	MSPE values of prediction models		
	SS	Starch	Protein
PCR	1.63	26.41	1.92
PLSR	0.68	28.03	0.98
LS-SVM	1.65	28.41	0.51
LS-SVM <sub>PC</sub>	0.86	27.69	1.52
LS-SVM <sub>PLS</sub>	0.65	27.62	0.62

## Final thoughts

- These results suggest to using LS-SVM after initial PCA and PLSA
- PLSR performed better than PCR and LS-SVM
- More work is needed before LS-SVM can be used routinely.

The background is a light blue gradient with a subtle grid of small squares. Overlaid on this are several dynamic, flowing lines in shades of blue and white, creating a sense of movement and depth. The lines curve and sweep across the frame, some appearing as thin, sharp streaks while others are more diffuse and ethereal.

**THANK YOU**