# Local vs. SVM–based stacking approach for predicting proportions of complex blends in food products

Beatriz Carrasco[1], Alberto Gil[1], José Ramón Quevedo[2], Elena Montañés[2], Guanghui Shen[3, 4], Lujia Han[3], Zengling Yang[3], Vincent Baeten[4], Juan Antonio Fernández Pierna[4]

1 Blendhub, 30169 San Ginés, Murcia, Spain
2 Artificial Intelligence Center, Universidad de Oviedo, Gijón, Spain
3 China Agricultural University (CAU), China
4 Walloon Agricultural Research Centre (CRA–W), Belgium

*Contact:* *bcarrasco@blendhub.com / j.fernandez@cra.wallonie.be*

## 1. Background and Motivation

Certain ingredients present in complex blends of food products play an important role in the processing industry for adding texture, enhancing tasting, etc. Controlling their proportion is essential both to assess the legal requirements for the safety of the consumers and to avoid fraud in food labelling. The samples available were processed using NIR technology to obtain their spectra and were labelled with the proportion of ingredients with which they were manufactured. This work compares two different methods for predicting such proportions of ingredients in food additives from NIR spectra.

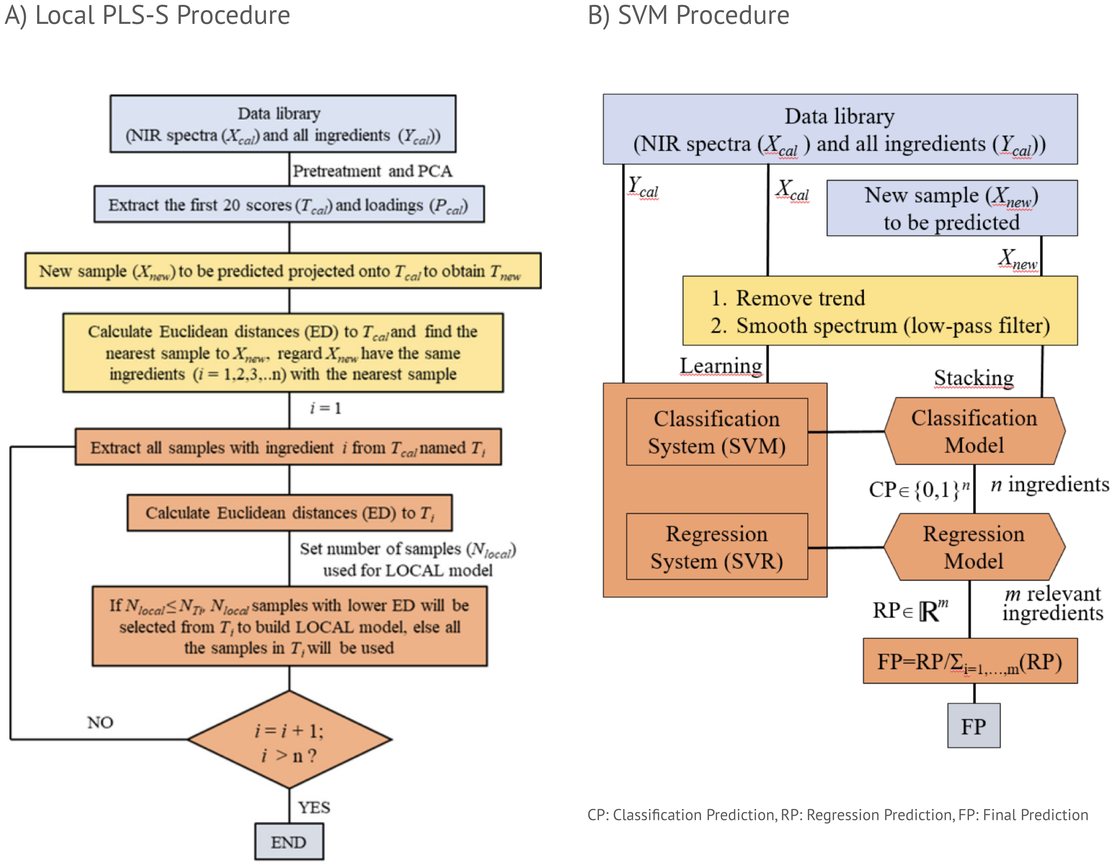
## 2. Methods and Material

A library built with 47377 spectra of finished products manufactured since 2014 to nowadays in Blendhub and recorded in a DS2500 Foss equipment (400-2500nm each 0.5nm) has been used. The spectra in the library are from 453 formulations and 331 ingredients gathered in 225 categories.

1000 samples were randomly extracted to validate externally the developed method.

A Local based Partial Least Squares (PLS) method and Support Vector Machines (SVM) have been used to predict the concentration of each ingredient in the blend.

## 3. Mathematical Procedure




CP: Classification Prediction, RP: Regression Prediction, FP: Final Prediction

Fig. 1 Schemes about how both methods work (SVM and LPLS-S).

## 4. Results and Analysis

The results have been analyzed obtaining the RMSEP of the predicted and the real values for each ingredient on the 1000 spectra used for validation. Usually, the results obtained by LPLS-S are better than results obtained by SVM for the prediction of the percentage of each ingredient. The categories where there were no enough samples to do the calibration have not been analyzed (in total 112 categories have been predicted).

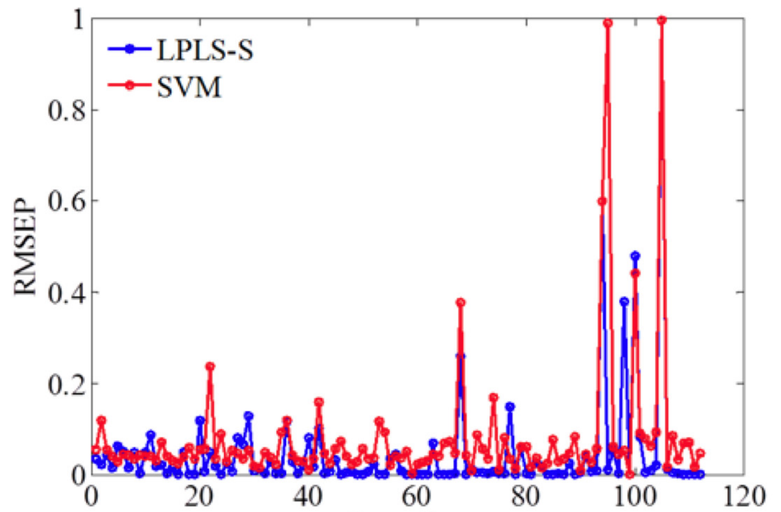



Fig.2 RMSEP results obtained for both methods for each ingredient where enough information for carrying out the analysis was available.

## 5. Conclusions

In summary, both methods (LPLS-S and SVM) can be successfully used to detect and predict the percentage of the ingredients in a blend. In general, LPLS-S has a better performance than SVM. It is important to remark the following aspects:

- LPLS-S does not need any pretreatment nor learning process
- The computing time required for SVM is around 2 hours while for LPLS-S is less than 4 minutes for the whole procedure.
- For LPLS-S, no need of individual calibration (and therefore no updates) for every ingredient.
- Misclassification increases when SVM is used. A cause could be that the optimization of the SVM parameters has been carried out globally. In a next step, an individual optimization for each individual ingredient will be carried out.
- These techniques open-up a chemometric alternative to study the composition of complex blends using NIRS. This composition analysis is very difficult or even impossible to carry out using other methods.

## 6. Acknowledgments

This research has been supported by Spanish Ministerio de Economía y Competitividad (grant TIN2015-65069-C2-2-R).

Blendhub

Wallonie recherche CRA-W

Universidad de Oviedo

中国農業大學 China Agricultural University

www.blendhub.com