# Galoá

#### **Book of Abstracts**

Use of NIRS, PLS and OPS Variable Selection to Predict Five Quality Parameters of Sweet Sorghum Juice used to Produce Bioethanol

> Marcelo Sena Cristiane Guimarães Camila Assis Maria Lúcia Simeone

#### **Track**

### 8. Agriculture and food

First generation biofuels represents a well established sustainable energy source, reducing pollutants emission in the atmosphere. A culture highly feasible to be used for the bioethanol generation in Brazil is the sweet sorghum, which is particularly promising as complementary for the diversification of sugarcane croplands due to its shorter harvest period. Sorghum breed programs have sought to develop new cultivars of higher bioethanol yield, generating a large number of juice samples that must be analysed and characterized before the fermentation. Thus, rapid and non-destructive analytical methods are required. In this work, multivariate calibration models were developed for determining five quality parameters of sorghum juice (Brix, saccharimeter reading/SR, reducing sugars/RS, Pol, apparent purity/Q) based on PLS and NIRS. All of these models were optimized by variable selection with Ordered Predictors Selection (OPS). OPS selects the most predictive variables by a systematic investigation of the PLS informative vectors (correlation coefficients, VIPscores, NAS, etc.) in a cross-validation process, leading to a great reduction in the number of variables.

500 juice samples were obtained from 275 recombinant inbred lines of sweet sorghum derived from two contrasting lines in relation to the quality and quantity of sugars. The five quality parameters were determined by appropriate reference methods. Juice samples were previously filtered in cotton and their spectra were recorded on a Büchi NIRFlex N-500 FT-NIR spectrometer, equipped with a transflectance accessory, from 10000 to 4000 cm-1, with 4 cm-1 steps. Data were handled with Matlab, PLS\_Toolbox and a homemade routine for OPS.

The analytical ranges for the parameters were 5.5-18.1°B (brix), 1.1-53.2°Z (SR), 1.2-5.2%

w/v (RS), 0.3-13.0% w/v (pol) and 9.8-83.0% w/v (Q). Samples were split in 333 for calibration and 167 for validation using the Kennard-Stone algorithm. The best models were obtained with the preprocessings first derivative, Savitzky-Golay smoothing and mean center. All the PLS models were built with 6 latent variables. Subsequently, the models were optimized by variable selection with OPS, reducing the number of wavenumbers used from 1501 (full spectra) to 50-130, depending on each parameter. All the RMSEC and RMSEP values for the OPS-PLS models were decreased. RMSEP were decreased from 0.4 to 0.3°B (brix), 3.2 to 2.1°Z (SR), 0.4 to 0.3% w/v (RS), 0.8 to 0.6% w/v (pol) and 5.8 to 5.3% w/v (Q). Correlation coefficients (r) for the reference versus predicted values were between 0.896 and 0.992. The randomness of the residuals was checked by appropriate statistical tests, assuring the linearity of the methods. The developed methods were validated by estimating figures of merit, such as trueness, precision, analytical sensitivity and bias. RPD (residual prediction deviation) was estimated between 2.4 and 6.2, attesting the good prediction ability of the models. The use of OPS allowed developing simpler, more interpretable and predictive multivariate calibration models. The NIRS methods are rapid, non-destructive and of low cost, being appropriate for replacing the more laborious reference methods, what is stressed considering the large number of samples analyzed.

## **Acknowledgements**

CAPES for a fellowship (C.A.), Reinaldo F. Teófilo (UFV, Brazil) for the OPS routine, and FAPEMIG for the financial support.

