



## Working Group on Statistical Learning Seminar

**Title:** Improving prediction from spectral data using reordered probabilistic principal component regression

**Speaker:** Suzy Whoriskey (University College Dublin)

**Date:** Mon 21st February 2022 at 12:00PM

**Location:** (See abstract)

**Abstract:** Prediction from high-dimensional spectral data can be a challenging statistical problem. Typically, the number of observations  $n$  is less than the number of variables  $p$ . In such settings, standard regression models fail. Dimension reduction techniques provide feasible alternatives, such as principal component regression (PCR) and partial least squares regression (PLSR), which link latent variables to the response. However, PCR does not necessarily find latent variables related to the response, while PLSR does not offer uncertainty quantification.

Here we propose reordered probabilistic PCR (rPPCR). This method extends probabilistic PCR (PPCR), which embeds PCR in a Gaussian latent variable framework. Moving to a probabilistic model allows for principled statistical inference and uncertainty quantification. Additionally, we seek to improve the prediction performance of PPCR by considering the correlation of its latent variables with the response of interest.

The rPPCR method is motivated and illustrated by predicting certain traits of in-

terest, firstly, from near-infrared spectra with  $p = 236$  wavelengths from 116 grain samples and, secondly, from mid-infrared spectra with  $p = 532$  wavelengths from 366 milk samples. Comparison of the rPPCR performance to that of PCR and PLSR demonstrates the utility of the proposed method.

\*\*\*\*\* The seminar will be held live in University College Dublin, room H0.12. \*\*\*\*\*  
The seminar will also be live-streamed over Zoom: <https://ucd-ie.zoom.us/j/68316324831>  
\*\*\*\*\*