**Using machine learning in chemometrics to check validity of existing data driven models for new data**

Olivier Paquet-Durand*\**, Supasuda Assawajaruwan, Bernd Hitzmann

*Process analytics and Cereal Science, University of Hohenheim, Garbenstr. 23, 70599 Stuttgart, Germany*

*\*Corresponding author:* [o.paquet-durand@uni-hohenheim.de](mailto:o.paquet-durand@uni-hohenheim.de)

**Highlights**

* Online validation of chemometric models
* Evaluation of fluorescence spectra
* Prediction of glucose, ethanol and biomass
* Supervision of cultivation of baker’s yeast cells

**1. Introduction**

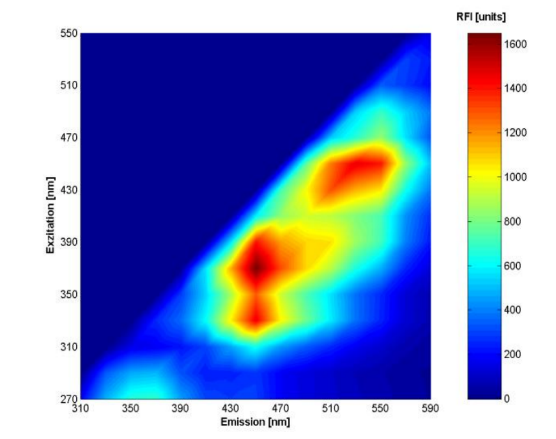
A big problem in chemometrics is, that small changes in process setup or equipment or just ageing of compounds might transform existing models useless [1]. Than the chemometric models must be recalibrated. The major catch is, that it is not obvious if a recalibration is necessary or not.

**2. Methods**

In this contribution we will demonstrate, how machine learning can be used to check existing chemometric models for validity with new data or changing process parameters. To do so, we used an autoencoder [2] which is trained during calibration of chemometric models with the same data and the same number of hidden neurons in the code layer as the chemometric model uses principal components. The basic idea is that when new data point are presented to the chemometric model, the autoencoder error is calculated as well. If this error is significantly higher than during calibration, the used chemometric model is no longer valid and need to be recalibrated. Otherwise a recalibration is not necessary. A 2D fluorescence spectrometer, which was applied in the study, is the BioView fluorescence spectrometer. The device is equipped with 15 different filters for excitation and emission wavelengths. The measurement of one spectrum using the BioView spectrometer has 120 fluorescence intensity variables of excitation and emission wavelength combinations; scattered light is not considered here [3].

**3. Results and discussion**

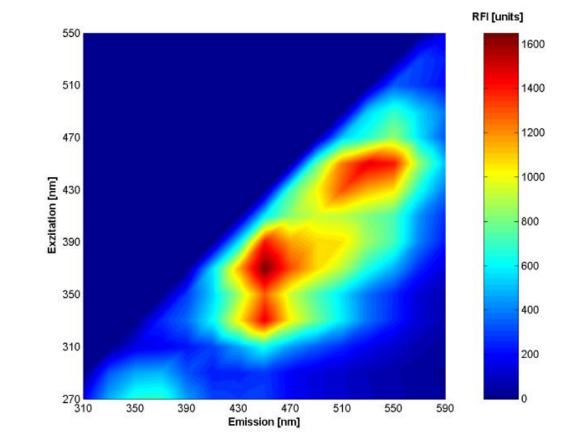
The online validation procedure is tested during a cultivation of baker’s yeast cells [4]. Here a network with 120 input neurons different hidden layers and 120 output layers are used. The ability of such an autoencoder for a validity check of data driven models will be demonstrated with the evaluation of fluorescence spectra. The data driven models are used for the prediction of glucose, ethanol and biomass. In Figure 1 the autoencoder system used here is present.



20

2

20



Code layer

**Figure 1.** The *autoencoder that is trained to predict the input as output. Within the encode or compression part, the number of neurons per layer progressively decreases. The bottleneck is the code layer with the least number of neurons. Within the decode or decompression part the number neurons progressively increase up to the same dimensions as the input.*

**4. Conclusions**

The online validation of data driven models is still not satisfactorily resolved [4]. In this contribution autoencoders are used to test the validity of a data driven chemometric model for the prediction of glucose, ethanol and biomass during the cultivation of baker’s yeast cells. Although the training of the network takes time, it is very worthwhile to obtain a tool, for the online validity check of the prediction capability of the chemometric model.

**References**

1. Dörte Solle, Bernd Hitzmann, Christoph Herwig, Manuel Pereira Remelhe, Sophia Ulonska, Lynn Wuerth, Adrian Prata, Thomas Steckenreiter, Between the poles of data-driven and mechanistic modeling for process operation, *Chemie Ingenieur Technik* 89,5(2017)542-561
2. Chicco, Davide; Sadowski, Peter; Baldi, Pierre; Deep autoencoder neural networks for gene ontology annotation predictions, *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics - ACM-BCB 2014*. page 533-540, doi:10.1145/2649387.2649442
3. Saskia Faassen, Bernd Hitzmann, Fluorescence Spectroscopy and Chemometric Modeling for Bioprocess Monitoring, *Sensors*, 15(2015)10271-10291
4. Supasuda Assawajaruwan, Philomena Eckard, Bernd Hitzmann, On-line monitoring of relevant fluorophores of yeast cultivations due to glucose addition during the diauxic growth, *Process Biochemistry*, 58(2017)51-59