Air vs. Water: Defining the Optimal Reference for NIR Spectroscopy of Aqueous Mixtures

The optimal choice of a spectral reference is important for building accurate calibration models. When water is used as a reference sample in the analysis of aqueous solutions (mixtures) by near-infrared (NIR) spectroscopy, the resulting spectra tend to be strongly distorted due to the presence of negative peaks (in place of the water signals). Although this makes the spectra less interpretable, their information content remains high, and their suitability for quantitative analysis is not impaired. Therefore, air and water can both be used as the reference in the quantitative analysis of liquid solutions. The present study aims to answer this question using an experimental approach. Two series of NIR-spectral measurements were obtained using air and water as references. The measurement was performed on the same set of specially designed samples of a ternary aqueous mixture of ethanol and methanol using a fiber-optic transflection probe. The references were compared on the basis of statistical validation of the obtained partial least-squares (PLS) regression models in each calibration case.

Water was found to be a preferred standard material in the studied case. Using the water sample spectrum as a reference measurement enables accurate prediction of the two alcohols under study without data preprocessing and with a lower number of latent variables, i.e. with a simpler, and hence, more reliable calibration model.

NIR-spectroscopy is widely used for quantitative analysis of solid and liquid samples. Solid samples are usually analyzed in the diffuse reflection mode with a white reflection standard as a reference material. In contrast, liquid solutions are predominantly measured in transmission (transflection) geometry in a cuvette or a using fiber optic probe. In this case, the air (no reference) and the water (or another solvent) are the two commonly used reference samples. Both substances obey the requirements to a standard material: they are widely available and reproducible in terms of their spectral properties. Air is a natural reference choice for qualitative analysis, where the spectrum is used to assign the spectral signals to fragments of a chemical structure that should be identified or confirmed. In the modern analytical practice, however, NIR-spectroscopy is one of the primary tools of quantitative analysis, particularly as a key method of in-line monitoring of component concentrations in course of a running process or in the field conditions.

The present investigation was performed using a designed set of 25 samples mixed in accordance with the diagonal experimental design for two-component mixtures [A. Bogomolov, *Diagonal designs for a multi-component calibration experiment*, Anal. Chim. Acta **951** (2017) 46-57]. Chemically pure methyl alcohol (99.9%) and ethyl alcohol (99.9%), as well as the distilled water were used for sample preparation. The exact component concentrations in water varied between 0% and 19.2% for both components.

NIR spectra of the samples in the range 930–1720 nm were obtained using a grating-based non-cooled NIR spectrometer (Broadcom, model Qneo AFBR-S20M2WV)) through a transflection fiber probe (art photonics GmbH, Germany) with an adjustable pathlength coupled with a stabilized tungsten-halogen light lource. In the present work the probe slit was set to 3.0 mm, which corresponds to the total optical pathlength of 6.0 mm, considering that the NIR-light passes the way twice due to the mirror reflection (Fig. 1). Two experimental series were analyzed where the air (no sample) or water spectra were chosen as references.

Figure 1. Tip of the Transflection probe used for the NIR-spectroscopic analysis of ethanol-methanol mixtures.

NIR-spectra of the mixture samples are presented in Figs. 2 and 3 obtained with the air and the water reference, respectively. In the calibration dataset, each sample was presented by three replicate spectra to account for the measurement reproducibility, when the probe was removed from the sample and reinserted.

As the raw spectra are dominated by the strong signal of water at 1450 nm (Figs. 2 and 3, top), their first or second derivatives can be used instead to emphasize the spectral differences of the vibration overtones of -OH, -CH3 and >CH2 groups of the alcohols (Figs. 2 and 3, bottom).

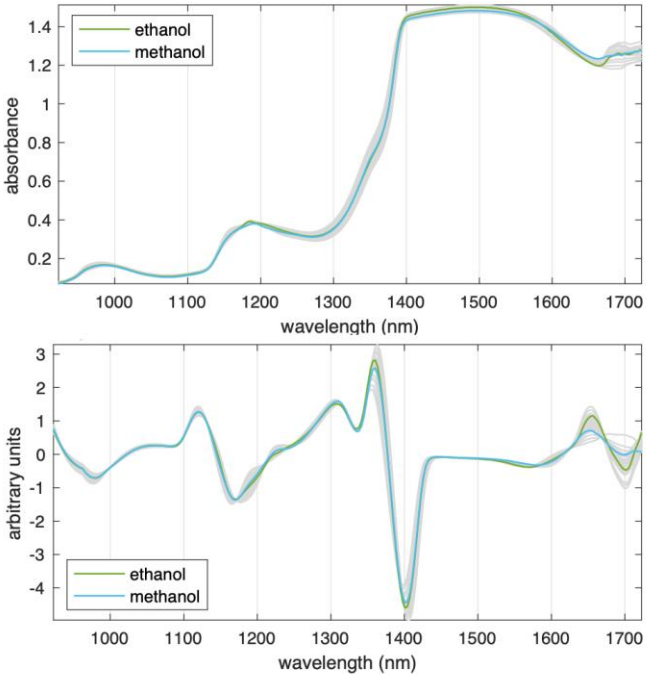

*Figure 2. NIR-spectra with **air** as a reference of 25 designed samples of aqueous ethanol-methanol mixtures: raw data (top) and first derivative by Savitzky-Golay (bottom).*

Despite the derivative transformation the main variance observe in the spectra is due to the difference in the water concentration in samples. Calibration models for ethanol and methanol were built using PLS-regression and spectral data in the range of 1100–1350 nm, as it has shown the best validation statistics in preliminary models.

The models were built using different preprocessing methods: raw data, first and second derivatives by Savitzky-Golay algorithm. The models were compared using the following validation statistics. The prediction accuracies were estimated using the root mean-square error (RMSE) and coefficient of determination (R^2) statistics segmented cross-validation (with segments formed by three replicate spectra of the same sample). The most accurate models are presented in Fig. 4 as predicted versus reference concentration plots with the model validation statistics. All calibration models were built with 4 latent variables (LVs). Although the presented models show very high accuracies in predicting the concentrations of both alcohols (the root mean-square error of approximately 0.3% over a wide concentration range), models built on NIR-spectra using the air reference and preprocessed using first-derivative transformation have slightly lower RMSE. However, models build on data obtained with the water reference sample do not require any preprocessing to achieve similarly high prediction accuracies for both alcohols.

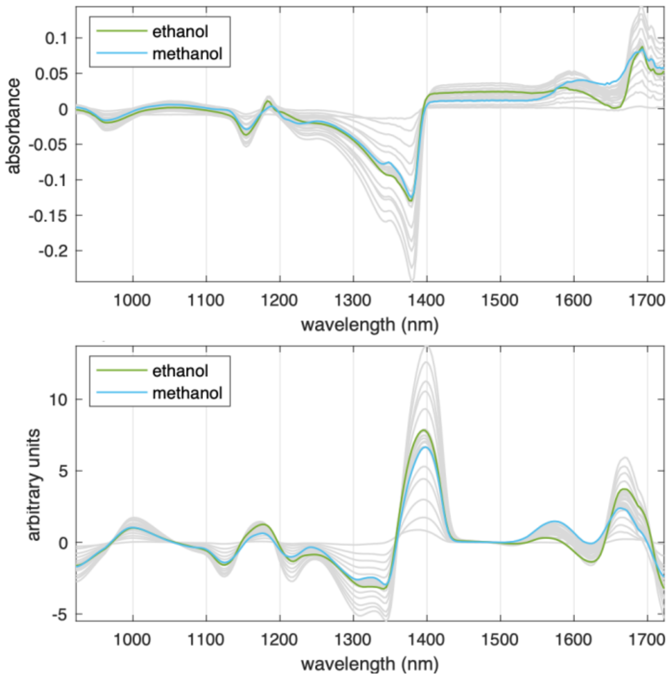

*Figure 3. NIR-spectra with **water** as a reference of 25 designed samples of aqueous ethanol-methanol mixtures: raw data (top) and first derivative by Savitzky-Golay (bottom).*

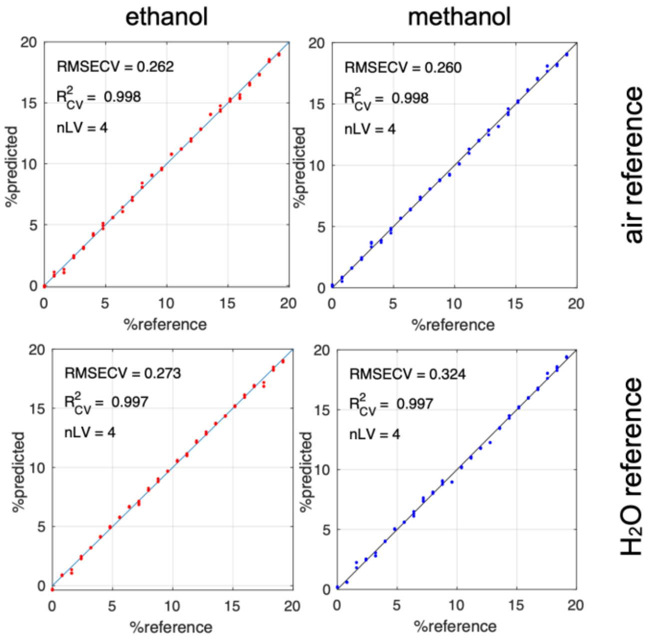

Figure 4. Cross-validation statistics of PLS calibration for methanol and ethanol using air (spectra were preprocessed by the first derivative transformation, Fig. 2, bottom) and water as references (no preprocessing, Fig. 3, top).

In conclusion, both air and water can be successfully used for reference analysis in the analysis of aqueous solutions. Air (no reference sample) can be preferred as a reference for its experimental simplicity. Using water as a reference simplifies data analysis by avoiding the spectral preprocessing step, which is required for air-based measurements.