Article

Calibration transfer for LED-based optical multisensor systems

Anastasiia Surkova, Andrey Bogomolov, Andrey Legin, and Dmitry Kirsanov

ACS Sens., **Just Accepted Manuscript** • DOI: 10.1021/acssensors.0c01018 • Publication Date (Web): 21 Jul 2020

Calibration transfer for LED-based optical multisensor systems

Anastasiia Surkova*[1,2], Andrey Bogomolov[2,3], Andrey Legin[1], Dmitry Kirsanov[1]

[1] Institute of Chemistry, St. Petersburg State University, Universitetskaya nab. 7-9, Mendeleev Center, 199034 St. Petersburg, Russia

[2] Samara State Technical University, Molodogvardeyskaya Street 244, 443100 Samara, Russia

[3] Endress+Hauser Conducta GmbH+Co. KG, Anthon-Huber-Strasse 20, 73430 Aalen, Germany

* Corresponding author: melenteva-anastasija@rambler.ru; +79277454932 ORCID: 0000-0001-6220-7822

ABSTRACT

Multivariate calibration transfer is widely used to expand the applicability of existing regression model to new analytical devices of the same or similar type.The present research proves the feasibility of calibration model transfer between a full-scale laboratory spectrometer and an optical multisensor system based on only four light-emitting diodes at different wavelengths. The model transfer between two multisensor systems of this kind has also been studied. Both possibilities were successfully performed without any significant loss of precision using a designed set of training and transfer samples. Direct standardization and slope and bias correction protocols for model transfer were tested and compared. The best model transfer between two optical multisensor systems was obtained using the direct standardization.

KEYWORDS: calibration transfer; optical multisensor system; optical spectroscopy; direct standardization; slope and bias correction.

Multisensor systems based on different physical principles is a rapidly growing field of research and development. It is a device composed of several (from a few to some tens) sensors with pronounced cross-sensitivity, i.e. sensitivity of each individual sensor towards several components of an analyzed sample simultaneously, in the absence of sharp selectivity to either of them. The lack of selectivity towards particular analytes is typically compensated by the application of modern methods of multivariate data analysis (chemometrics). Electrochemical (potentiometric and voltammetric)[1–3] and optical sensors[4–7] are the most popular platforms to construct multisensor arrays for the analysis of gaseous and liquid samples. Some other unit types, such as surface

acoustic wave and mass-sensitive sensors, can also be employed in multisensor systems[8,9].

Optical multisensor systems (OMS) can be based on a set of light-emitting diodes (LEDs) with emission maxima at preliminary optimized wavelengths[10] combined with a photometric detector. The multisensor concept allows for a significant price, size and weight reduction of the sensing device and, unlike full-size spectrometers, can be easily employed for on-site and on-line applications. Recent examples include the evaluation of white grape ripeness using multisensor system with four LEDs[11], classification of tea types using LED-induced fluorescence[12], and recognition of tumors in kidney tissues[10].

The output data of OMS includes only a few bright-band variables corresponding to working wavelength regions, which can potentially overlap and their efficient modelling may require some modification of existing mathematical approaches. Building multivariate calibration models, such as partial least squares (PLS) regression[13], is routinely performed at different development stages of spectroscopic analyzers of any kind. Calibration model relates the system response to a measured property of the sample, e.g. to the concentration of an analyte. Robust calibration should be built on a representative training set of samples analyzed by a reference analytical method. As the real-life applications usually involve complex mixtures affected by numerous sources of compositional variance, a significant number of designed[14,15] or historical[16] samples is required to train the model. This makes the multivariate model one of the most expensive elements of any multisensor system. The necessary training samples with sufficiently large variability in the composition can often be designed under controlled laboratory conditions and measured with a high-precision full-scale spectrometers, thus resulting in a good-quality calibration model. This is not the case for analyzers working in the industrial conditions where the variability in sample composition is governed by the technological process and its intentional changes can be impossible. Therefore, the model transfer from a full-scale laboratory spectrometer seems to be an attractive option for calibration of industrial or in-field multisensor systems.

Another typical problem faced by the developers of OMS is a difficulty (or impossibility) to construct two identical systems that is mainly related to the fact that OMS is conceptually introduced as a budget analyzer. Therefore, it tends to use poorly standardizable materials, such as light fibers, 3D-printed parts, stamped optics etc.[17],

and to minimize standardization efforts in production. The lack of hardware standardization can be compensated mathematically at the data analysis stage, using the same principles of model transfer often called instrument standardization in that case[18]. The literature describes a number of methods for calibration model transfer[18–23] forming two basic groups. Some methods apply mathematical transformations to the raw spectral data, while others perform a correction of the very modelling result: the vector of regression coefficients or predicted values.

The purpose of this study was to explore both issues of the model transfer in multisensor system development: from a full-scale laboratory spectrometer to a multisensor system and between two optical systems of the same kind. To the best of our knowledge there are no other transfer methods available that would allow handling the tricky situation of dramatically different number of measuring channels in the standardized instruments (2500 wavelengths in the spectrometer vs 4 LED channels in OMS). For example, the classical piece-wise direct standardization[21] is not applicable in the case of spectrometer-to-OMS transfer, since the number of variables in their response vectors is different. In the present study we compare two different types of model transfer that can address this situation. Direct standardization (DS)[20] applies mathematical transformations to the raw instrument responses, while “slope and bias correction” (SBC)[18] performs a correction of the model (predicted values). DS was recently successfully applied[22] to a model transfer between two different instruments: X-ray fluorescence spectrometer and UV-Vis spectrometer, and also between two sensor arrays based on different measuring principles[23]. The SBC method has also demonstrated a good performance even in complicated transfer case. An example can be found in[16], where the models for fat and total protein determination in milk were transferred between two spectrometers by means of a designed set of milk samples. The above described model transfer methods were tested on the developed series of standard samples following the diagonal design[14] and on the real data obtained in the spectroscopic quantification of ammonium nitrogen in surface and tap water samples. General feasibility of the model transfer in the case of OMS and the prediction performance of different techniques is discussed.

MATERIALS AND METHODS

Instruments (Sensor construction and data acquisition). Full-scale measurements were performed using the compact UV/Vis/NIR spectrometer Q-mini by RGB Photonics GmbH (Kelheim, Germany). The spectra were obtained in the wavelength region 180–1008 nm with 0.35 nm step in the 10-mm glass cuvette through fiber optic connectors. Distilled water was used as a reference measurement for absorbance calculation.

OMS measurements were performed with two LED-based optical multisensor system prototypes (Fig. 1) by Endress+Hauser Liquid Analysis GmbH+Co (Aalen, Germany). These instruments are further encoded as “OMS-1” and “OMS-2”. The optical multisensor system consists of four LEDs with the working wavelengths at 280, 340, 600, and 860 nm (emission maxima). These wavelengths were chosen to provide the signals in the whole spectral range from UV to short-wave NIR. Spectra of four LEDs from OMS-1 acquired on Q-mini are shown on Fig. 1S. For the measurement, the probe head of the sensor (Fig. 1) was inserted into a plastic cuvette with 1.5 mL of the sample solution. The light alternately emitted by the LEDs working in a pulse mode was collected by optical fibers and transferred to the solution filling the 5-mm gap of the transflection probe. The incident light after passing through the solution was reflected back from the mirror on the opposite side of the gap and collected by the detector fibers. The resulting multisensor response consisted therefore of four values indicating signal intensity at each of the four wavelengths. The raw detected values at each wavelength were stabilized by internal reference measurements and dark detector signal correction. The raw signals in mV were employed for all the modelling and plotting without calculations of absorbance. The output data of OMSs were absolute values in arbitrary units.

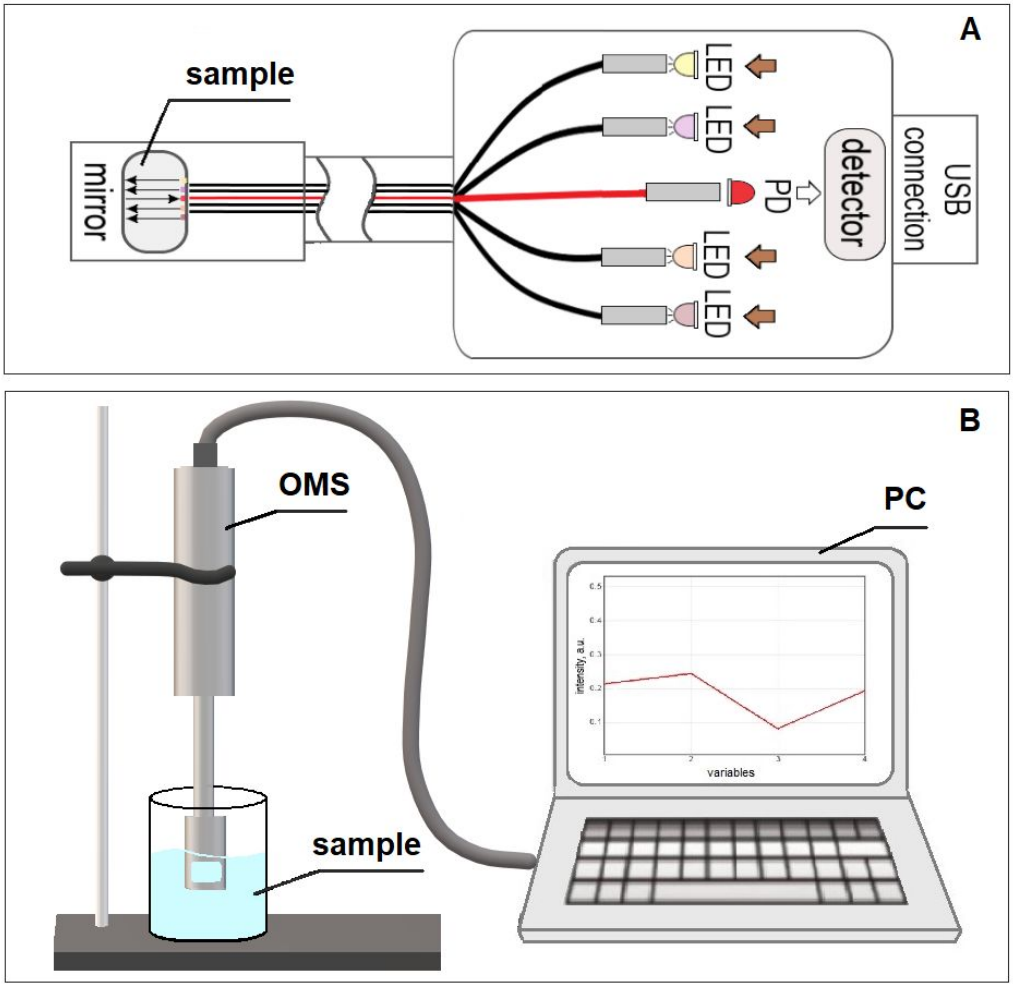


Figure 1. A LED-based optical multisensor system: (A) schematic drawing of the main construction parts and (B) measurement setup.

The measurements on both instruments were carried out in the plain absorption mode. To test the reproducibility of OMS a preliminary study was performed on the standard samples of water. The relative standard deviation between the replicated measurements does not exceed 0.83%. The reproducibility plot for three measurements of water and related statistics for the OMS-1 is shown on Fig. 2S and in Table 1S.

In order to demonstrate the applicability of the proposed methodology to real samples UV/Vis spectra of surface and tap water samples were acquired with UV/Vis fiber-optic spectrometer AvaSpec-ULS2048CL-EVO (Avantes, Apeldoorn, the Netherlands) and standard procedure for quantification of ammonium nitrogen in aqueous samples was applied. The measurements were performed in 1 cm glass cuvette with water as a reference at room temperature. The halogen lamp AvaLight-HAL-Mini (Avantes, Apeldoorn, the Netherlands) was used as a light source. All

spectra were recorded in triplicate in the spectral range 365–425 nm and then they were averaged.

Simulated OMS. The performance of the suggested transfer procedures with the real data was assessed with four simulated OMSs. Four channels of each simulated OMS were modelled as arbitrarily selected intervals with a width of 10 or 20 nm from full spectra of AvaSpec spectrometer. These spectral intervals were averaged and multiplied by random number from 0 to 1. The width of the intervals were allowed to overlap partially.

Samples. To show the fundamental possibility of using high-precision models obtained on a conventional spectrometer for a LED-based OMS, a possibly simple and non-numerous set of calibration and transfer standard samples was proposed. Following the principle of simplicity, 25 samples were prepared by mixing aqueous solutions of cobalt(II) and copper(II) nitrate in different proportions in accordance with the scheme of the diagonal experimental design[14] (Fig. 3S). The samples in this design are placed along the diagonals of the concentration space following the principle of Latin square, so that each concentration level is represented by only one sample. This scheme provides a uniform coverage of each component's concentration ranges in many levels. At the same time, the correlation of component concentrations is minimized. The concentrations of both $Co(NO_3)_2$ and $Cu(NO_3)_2$ ranged from 10^{-2} to 10^{-1} M (Table 2S) to meet optical sensitivity of OMS channels of the OMS.

The metal salts were chosen to have absorption bands in the working spectral regions of the LEDs used in OMS. Absorption spectrum of the copper(II) nitrate has a broad band in the wavelength region 700–900 nm with a maximum around 800 nm that covers LED emission bands of two sensor channels: 600 nm (peak tail) and 860 nm (close to peak maximum). Both copper(II) and cobalt(II) nitrates strongly absorb in the UV and are therefore detectable by the 280-nm channel of the system. The broad band of cobalt(II) nitrate in the Vis region (400–600 nm with a maximum around 540 nm) does not interact with the sensor channels with an exception of the very tail of the 600-nm source. Using a mixture of copper(II) and cobalt(II) nitrates instead of pure solutions of the salts complicates the determination of substances due to signal interference. Another complication of the chosen mixture is that the LED-channel at 340 nm has low sensitivity to both analytes.

Therefore, the designed sample set simulates two real-life complications of mixture analysis by optical multisensor systems: component interference that can be higher and lower at different channels and a big difference in channel sensitivity of the system.

To test the performance of the suggested inter-instrument transfer procedures with the real data six water samples were collected from tap, river, lake and pond in Saint Petersburg region.

In order to determine the quantitative content of ammonium nitrogen in water samples the standard photometric method employing Nessler's reagent was applied. Calibration set consisted from 16 samples that were prepared as follows: 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, 2.75, and 3 ml of ammonium chloride with a mass concentration of 0.05 mg/cm^3 were diluted to 50 ml in volumetric flasks and brought to the mark with distilled water. Thus, the mass concentrations of ammonium ions in the prepared calibration solutions were 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, 1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, 2.75, and 3 mg/dm^3, respectively. 1 ml of sodium-potassium tartrate was added to each sample of calibration and test sets, mixed, and then 1 ml of Nessler's reagent was added and mixed again. The samples were analyzed after 10 minutes of conditioning.

Model transfer by direct standardization. Direct standardization protocol[19] was used to transfer the models from the spectrometer to an OMS or between two sensor systems. The transfer is always performed between the primary device (e.g. spectrometer used to acquire the calibration data) and the secondary one (an OMS intended for the application of transferred model) called "master" and "slave", respectively. The transfer is based on a special set of standard samples that are measured with both "master" and "slave" instruments (transfer set). The number of samples in the transfer set should be as small as possible, however it has to provide a sufficient coverage of the concentrations ranges of target analytes and has at least three levels to address possible non-linearity. In accordance with these these requirements, five transfer samples were selected from the diagonal design scheme including four samples at the corners and one sample in the center were chosen as representative (Fig. 3S, IDs ***1***, ***2***, ***13***, ***24*** and ***25*** from Table 2S). In the case of ammonium nitrogen determination in water, 5 samples with mass concentration 0, 0.3, 1, 2, and 3 mg/dm^3 were chosen as a transfer set.

The transformation matrix *F* is calculated using the following equation:

$$F = X_{tr\_master}{}^{+} X_{tr\_slave} \quad (1)$$

where X_{tr_master} and X_{tr_slave} are the spectral datasets acquired in transfer samples using the "master" and the "slave" instruments, respectively.

Then, the matrix *F* is used to convert the "slave" data into the "master" format (Eq. (2)).

$$X_{cor} = X_{slave} \cdot F \quad (2)$$

where X_{cor} is the corrected data from the "slave" converted into the format of the "master", X_{slave} is the original "slave" data excluding 5 transfer samples.

The reproduced data are used for prediction on the "slave" data using the "master" model (Eq. (3)).

$$\hat{y} = X_{cor} \cdot b_{master} \quad (3)$$

where $\hat{y}$ is the transferred prediction and b_{master} is the vector of regression coefficients of calibration model built with "master" data *X*.

Model transfer by slope and bias correction. The slope and bias correction method makes use of the mathematical transformation of the predicted values. The only prerequisite for this transformation is the linear dependence between the reference and predicted values for both "master" and "slave" instruments. Then the correction of predicted values can be performed using two coefficients describing the deviation of the predicted-versus-reference line slope and bias of the "slave" from 1 and 0, respectively. Although mathematically this correction can be done using only two experimental points, a larger number is usually required to compensate for the random error. The predicted values $\hat{y}$ for the new measurements are corrected in accordance with the following equation:

$$y_{cor} = (y_{ref} + \Delta b_0) \times \Delta b_1 \quad (4)$$

where y_{ref} and y_{cor} are raw and corrected values for the secondary instruments, respectively; Δb_0 and Δb_1 are SBC-coefficients, respectively.

Data analysis. In order to build the quantitative prediction models different regression method were tested. PLS regression is the most widely applied multivariate calibration algorithm in optical spectroscopy[13]. At the same time, as the number of

variables (measuring channels) in multisensor data is rather small it seems reasonable to use multiple linear regression (MLR)[24]. The main advantages of MLR are its simplicity and small computational burden. The MLR does not use factor projection and solves the regression equation (5) directly from X and y:

$$b = (X^T X)^{-1} X^T y \quad (5)$$

where b is the regression coefficient vector.

The model performance is characterized by the root mean-square error (RMSE, Eq. (6)) of calibration (RMSEC) and cross-validation (RMSECV) and by the respective coefficients of determination R^2 (Eq. (7)).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{k}\left(\hat{y}_i - y_i\right)^2}{k}} \quad (6)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{k}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{k}(y_i - \bar{y})^2}, \quad \text{where} \quad \bar{y} = \frac{\sum_{i=1}^{k} y_i}{k} \quad (7)$$

where y_i and $\hat{y}_i$ are known and predicted values, k – is the number of samples in the validation set (for CV k is the number of samples).

Matrix calculations for model transfer were performed in MATLAB R2008b (The MathWorks Inc., Natick, MA, USA). PLS and MLR models were built in web-based chemometrics software TPT-cloud (tptcloud.com) by Global Modelling (Germany) and Mestrelab Research (Spain).

RESULTS AND DISCUSSION

Individual regression models for spectrometer and OMS. Absorbance spectra obtained with Q-mini spectrometer and the data from OMS for 20 studied samples (without 5 samples of the transfer set) are shown in Fig. 2A, B, and C. The data of the three selected samples are highlighted to visualize the effects of compositional differences. To simplify the data comparison between the spectrometer and OMS, the spectra in Fig. 2A, D and E were truncated to the interval of 280–860 nm.

In general, the spectra below 360 nm (Fig. 2A) reflect spectral properties of the analytes (section “Samples”).

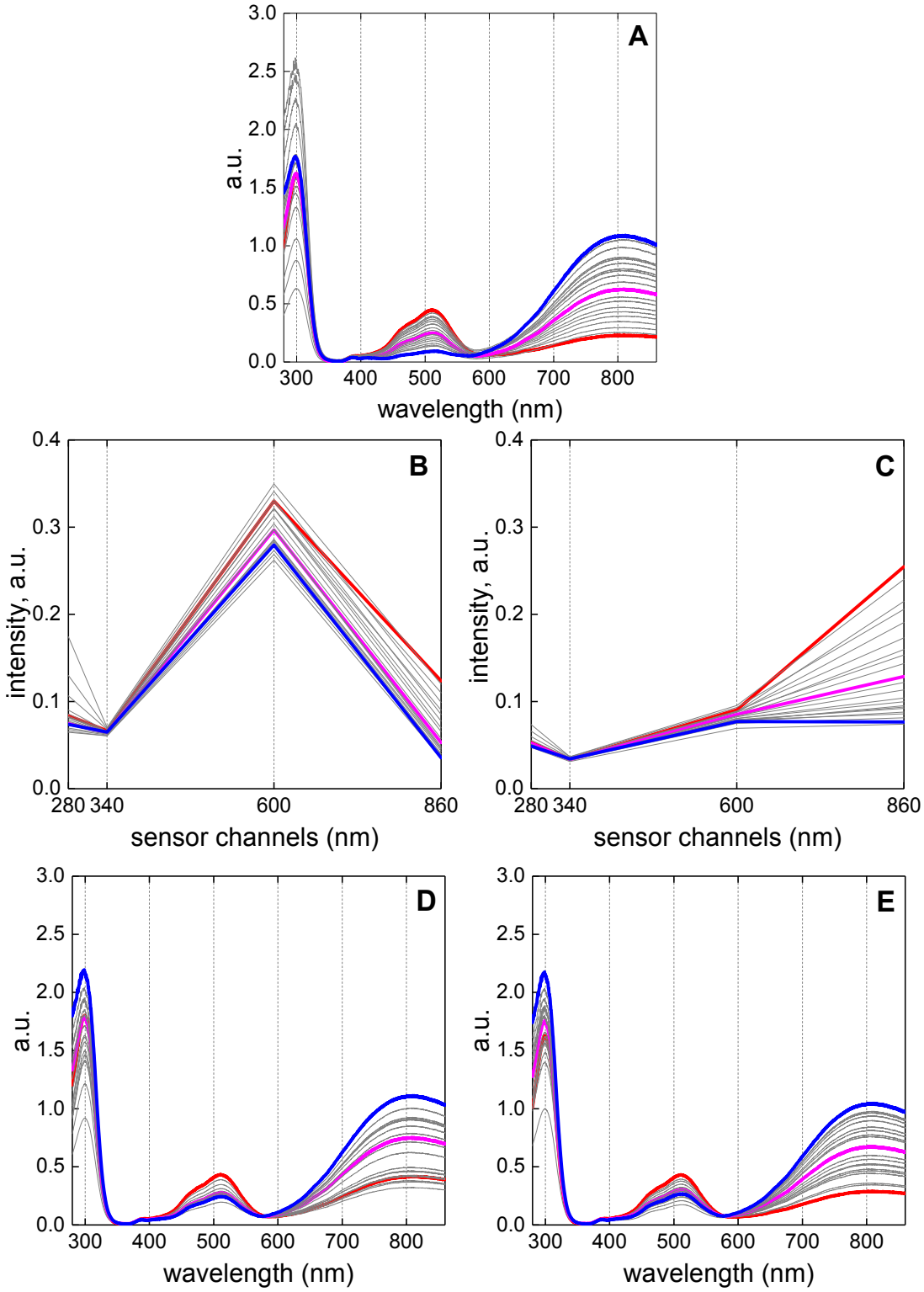


Figure 2. Data of 20 samples (without five transfer samples) obtained with: (A) Q-mini spectrometer (in the range 280–860 nm), (B) OMS-1 and (C) OMS-2; pseudo-spectra

X_{cor} reproduced from the multisensor data: (D) OMS-1 new; (E) OMS-2 new. Sample with the highest content of cobalt and low copper is marked red (IDs ***14*** from Table 2S); sample with the highest content of copper and the lowest cobalt is marked blue (IDs ***23*** from Table 2S); sample with the same content of c copper and cobalt is marked purple (IDs ***14*** from Table 2S).

Fig. 2B and C represent the measurements of the studied OMS: OMS-1 and OMS-2. Although all multisensor measurements were performed on the same sample set, their shapes (intensity ratios at different wavelengths) are dramatically different. This difference reflects a very poor standardization of the serially produced analyzers. In the present case it can be related to the application of optical fibers used to deliver the light to/from the sample. The quantity of light photons from the light source entering the fiber is strongly dependent on the precise adjustment of the optics (Fig. 1), which represents a big technical challenge that was not properly addressed yet. None of the multisensor intensity patterns agrees with the spectra in Fig. 2A (higher spectral absorbances correspond to lower raw signal detected by OMS). The brightness of LED pulses that can be compared using maximum signal intensity in the given series depends on many factors including their constructional differences and operating settings. Since no standardization was applied to equalize the LED channel brightness in any sense, no correspondence to the spectral absorbance is generally expected.

The detected signal variance at individual LED-channels correlates (in addition to the commonly expected dependence on the analyte absorbance in the region) with the channel general brightness; therefore, brighter channels are more informative in terms of the quantitative analysis. Thus, the 860-nm channels is respectively weak in OMS-1 (Fig. 2B), while 600-nm channel have lower brightness in OMS-2 (Fig. 2C). The channel at 340 nm has generally low informativeness, because it covers the region of low absorbances. Similarly, low spectral absorbance at 600 nm is compensated by the high channel brightness in OMS-1 (Fig. 2B). It can be therefore concluded that all investigated OMS have quality issues, related to the individual channel performance, and may also result in the data of very different quality. This drawback represent an additional (besides the low variable number) challenge for the model transfer in the studied cases.

The performances of two regression methods: PLS and MLR were compared in building the calibration models for responses of Q-mini spectrometer and two LED-based analyzers. The choice of the classical MLR algorithm as an alternative approach

is related to the known fact that for a small number of optimized (non-correlating) variables MLR can be even better than more sophisticated projection-based techniques, such as PLS[25]. In MLR calculations with Q-mini data only four spectral intensities at the wavelengths of OMS LED emission maxima (280, 340, 600 and 860 nm) were used.

The regression statistics is presented in Table 1. Since the main idea of this study is to demonstrate the fundamental possibility of model transfer from the spectrometer to the OMS, only cross-validation was done to compare the model performances. Application of an independent test set could change the observed statistics slightly, but it would not affect the overall demonstrated possibility of such transfer. The complexity of all PLS models assume the involvement of just 2 latent variables (LVs). At such a low model complexity 20 designed samples are sufficient for the future reliability of prediction. PLS-models for spectral data have good prediction performance both for copper and cobalt salts (Table 1 and Table 3S). The root-mean square errors of cross-validation (RMSECV) are 0.003–0.005 M for $Co(NO_3)_2$ and $Cu(NO_3)_2$ content and the respective R^2-values are 0.97–0.99. However, the models for $Co(NO_3)_2$ content built using the LED-based multisensor data are noticeably worse (Table 3S). This can be explained by the absence of a LED-channel in the region of cobalt(II) nitrate main absorbance (Fig. 2A). Therefore, the model transfer was only performed for the quantification of copper(II).

PLS-models for the multisensor data have RMSECV in the range 0.005–0.007 M of copper(II) nitrate content at R^2-values 0.93–0.95. This is slightly lower than those for the full-spectral models obtained with Q-mini data, however, still indicates the capability of accurate quantitative analysis with OMS.

The performance of MLR regression was superior to that of PLS for all studied instruments. It can be explained by the fact that the MLR works better when there are only a few explanatory variables in *X* to be used for predicting *y*. Moreover, the redundant information from the whole spectrum is not employed for modelling in case of Q-mini.

Table 1. PLS and MLR statistics for prediction of copper content. The models constructed using the instrumental responses in 20 samples. All PLS models are based on 2 LV.

Equipment	Calibration	CV

	RMSE, M	R^2	RMSE, M	R^2
PLS				
Qmini	0.004	0.97	0.005	0.97
Qmini [a]	0.004	0.97	0.005	0.97
OMS-1	0.006	0.95	0.007	0.93
OMS-2	0.004	0.97	0.005	0.95
MLR				
Qmini [a]	0.004	0.98	0.005	0.95
OMS-1	0.004	0.97	0.005	0.95
OMS-2	0.003	0.98	0.005	0.95

[a] the model was built with spectral signals for four wavelengths only: 280, 340, 600, 860 nm

The cross-correlation of both OMS with Q-mini spectrometer at the LEDs wavelengths and between two OMS is shown in the Fig. 4S, Fig. 5S and Fig. 6S, respectively. It can be seen that channels are correlated, but not always linearly. Nevertheless, the application of model transfer methods is possible.

Model transfer between OMS. In this section, model transfer between two multisensor systems is studied using two different methods. OMS-1 was always used as a "master" instrument and OMS-2 as a "slave".

The correction matrix for DS was calculated using the data acquired for five transfer samples (section "Model transfer by direct standardization"). The "slave" data for 20 other samples were corrected. Fig. 3A presents an example of data correction result for a chosen sample (IDs ***3***, Table 2S) acquired at OMS-1 (blue line) and reproduced using DS from OMS-2 response (green dotted line). It is clearly seen that the shape of the measured multisensor signal for "master" OMS-1 instrument is adequately restored for the "slave" OMS-2 data, although some intensity differences are present, mainly in the 860-nm channel. Note that OMS-1 has lower brightness of this channel compared to OMS-2 (Fig. 2B).

In accordance with DS model transfer, all corrected responses of OMS-2 were employed for the prediction of copper(II) content using a PLS (Fig. 3B) or MLR calibration model built with OMS-1 data. The statistical method comparison presented in Table 2 shows that the performance of the transferred model became slightly worse than the "master" model (see the corresponding data in Table 1), but the prediction error is still reasonably low (RMSEP=0.006 M).

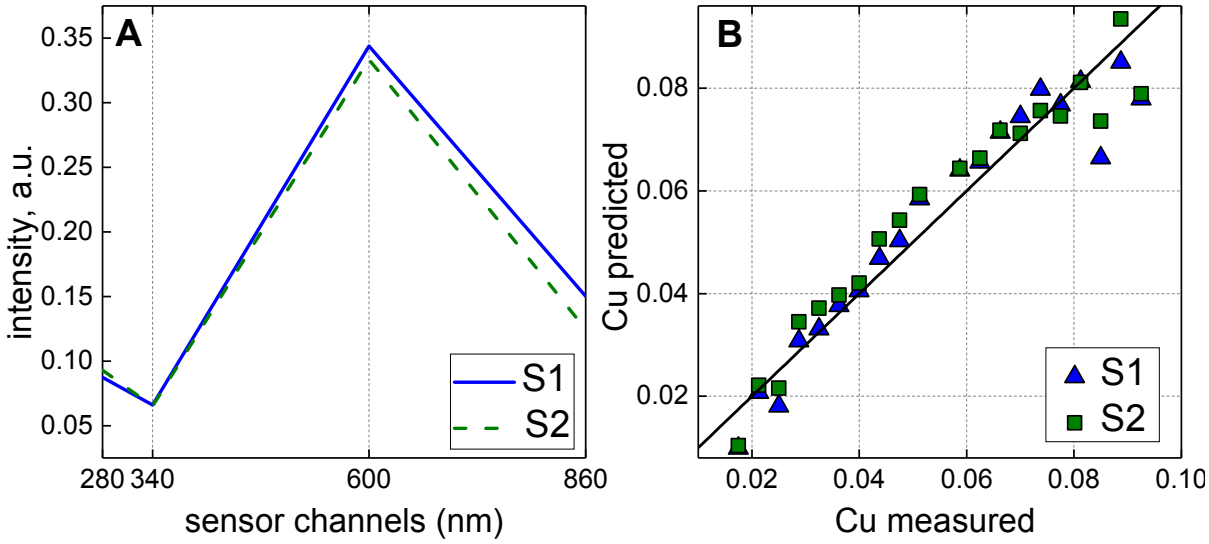


Figure 3. (A) OMS data of the same sample (IDs ***3*** from Table 2S) acquired with OMS-1 ‘‘master” (blue line) and corrected data of OMS-2 using DS (green dotted line); (B) predicted versus measured values for PLS-model for $Cu(NO_3)_2$. Values from ‘‘master” instrument (OMS-1) and corrected data of OMS-2 predicted on ‘master” model are designated by blue triangles and green squares, respectively.

Table 2. The results of model transfer between OMS for 20 samples. All PLS models are based on 2 LV.

Method of model transfer	Prediction	
	RMSE, M	R^2
PLS		
DS	0.006	0.93
SBC [a]	0.007	0.91
SBC [b]	0.013	0.68
MLR		
DS	0.004	0.97
SBC [a]	0.006	0.93
SBC [b]	0.008	0.87

[a] 20 samples or [b] 5 samples were used for model transfer by SBC

To perform SBC the dataset consisted of 25 samples from OMS-2 was predicted using OMS-1 model. For the more appropriate comparison with DS, the same five transfer samples (section “Model transfer by direct standardization”) were applied to find the slope and bias coefficients. Then, the rest of the samples (20) were corrected by means of these coefficients in accordance with equation (4) and were

employed for the assessment of prediction accuracy. The result was significantly worse both in PLS and MLR compared to DS (RMSEP=0.013 and 0.008 for PLS and MLR, respectively, Table 2). It can be explained by the fact that in this case the random error is large and in order to set the correct direction of the line, more samples are required for transfer.

To evaluate the effect of the number of the transfer samples the model transfer by SBC method was also performed with 20 samples (except for 5 samples that were previously used for transfer). Fig. 4 shows the difference in predicted content of copper(II) nitrate content using the ‘‘master” MLR-model for 20 calibration samples for OMS-1 and OMS-2 data before correction. It is clearly seen that predicted values for the ‘‘slave” instrument spectra are gradually shifted upwards compared to the „master” values spread around the quadrant diagonal, as expected. Since the predicted-versus-reference dependence for OMS-2 keep a linear trend, SBC method can be used (Eq. (4)). In this instance RMSEP = 0.006 was found for the MLR model, which is comparable with the results obtained using DS. The result of PLS regression of this type of transfer is shown in Fig. 7S. Again, the respective prediction statistics is somewhat worse than in both MLR and PLS ‘‘master” models (Table 2), but the observed model deterioration is acceptable.

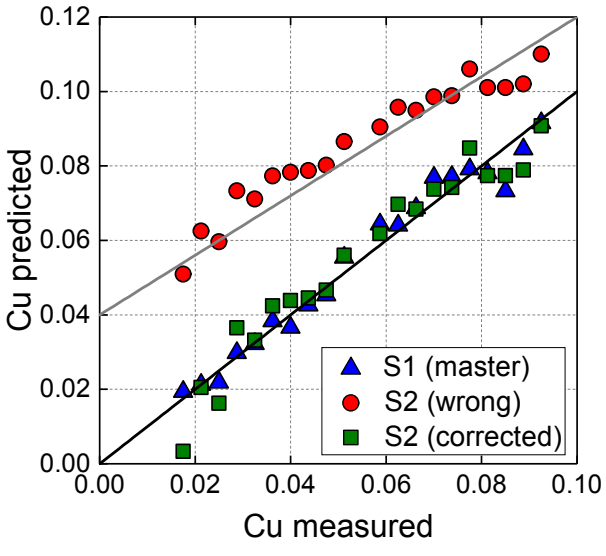


Figure 4. MLR-model transfer between OMS-1 and OMS-2 using SBC with 20 transfer samples: predicted versus measured values for $Cu(NO_3)_2$ model, where values from ‘‘master” instrument (OMS-1), wrong and corrected values from ‘‘slave” instrument (OMS-2) are designated by blue triangles, red circles and green squares, respectively. Gray line is the first quadrant diagonal and black line is the regression line.

The obtained experimental results indicate that DS procedure based on MLR provides more accurate prediction results compared to SBC method that requires more transfer samples to obtain a reliable model. Nevertheless, both DS and SBC methods can be applied for model transfer between OMS, because the obtained results are comparable. As expected, the model transfer between OMS using MLR works better than PLS. This observation can be explained by the simplicity of MLR, which makes it more suitable in the case of a small number of variables.

Model transfer from spectrometer to OMS. The datasets from the Q-mini spectrometer and an OMS have a different number of variables (2500 and 4, respectively). Nevertheless, the mathematics behind the DS allows handling this type of conversion. As it is shown in the Fig. 2D and E the pseudo-spectra X_{cor} reproduced from the multisensor data are basically similar to the spectra obtained with Q-mini.

Converted spectra of OMS were employed for prediction of $Cu(NO_3)_2$ content using PLS model built for Q-mini. The results of model transfer by DS show comparable behavior with satisfactory accuracies for both OMS-1 and OMS-2 (Table 3). The “native” RMSECV values were 0.007 and 0.005 M of copper for OMS-1 and OMS-2, respectively, and 0.005 M of copper for Q-mini. After DS the RMSEP became 0.006 M of copper for both sensors. The results of DS may indicate that model transfer works with an insignificant loss in precision and even with a slight improvement in the case of OMS-1. The results of the MLR and PLS regressions are almost the same, however, the PLS is slightly better. Since the sensor channels (280, 340, 600, 860 nm) were not optimized for this particular analytical task, PLS works better on full-spectra data than the MLR on four seleced spectral variables. PLS-predicted versus measured plots for OMS-1 and OMS-2 devices after the model transfer are presented in Fig. 5A and B.

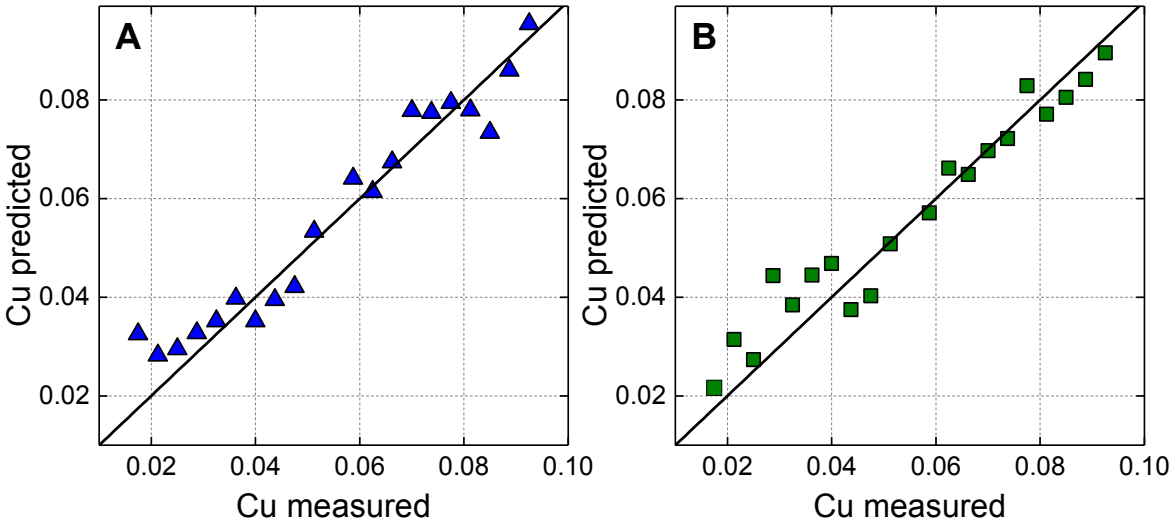


Figure 5. Model transfer results from spectrometer to OMSs by DS. PLS-predicted versus measured values of copper(II) nitrate content for: (A) OMS-1, (B) OMS-2.

Table 3. Results of model transfer from the spectrometer to OMSs (original models for Q-mini and OMSs are shown in Table 1). The concentration range of copper content content is 10^{-2}–10^{-1} M. All PLS models are based on 2 LV.

Method of model transfer	"slave" instrument	Prediction	
		RMSE, M	R^2
PLS			
DS	OMS-1	0.006	0.94
	OMS-2	0.006	0.93
SBC [a]	OMS-1	0.006	0.92
	OMS-2	0.006	0.93
MLR			
DS [a]	OMS-1	0.006	0.92
	OMS-2	0.007	0.91
SBC [a]	OMS-1	0.008	0.88
	OMS-2	0.008	0.88

[a] SBC for PLS and MLR models and DS for MLR-models were performed on 280, 340, 600, 860 nm

As shown in the section "Model transfer between OMS", the use of a small number of samples for model transfer by SBC yields insufficient accuracy. Therefore, SBC from the spectrometer to OMS was tested with 20 samples. SBC method works as good as DS for PLS models for both OMS and slightly worse for MLR.

In practice it can be sometimes necessary to address an inverse problem – the use of the model built with OMS using historical samples measured on-line in some

industrial process together with the data obtained from the full-scale laboratory spectrometer. In order to address this case the model transfer from the OMS to the spectrometer was also studied. The models for OMS were applied to the transferred data from the spectrometer. Table 4S shows the results of this type of transfer. As one would expect, the accuracy of the transferred models in this case is deteriorated, however, a reasonable accuracy is still observed.

Determination of ammonium nitrogen in water. The practical application of the proposed model transfer procedure was performed for the determination the ammonium nitrogen in water. Total ammonia nitrogen consisted of ammonium ion and unionized ammonia is one of the important water quality parameters. Therefore, its content is regulated by environmental standards and it is a subject to mandatory control. The standard procedure for quantification of ammonium nitrogen is based on spectrophotometry of colored sollutions after reaction with Nessler's reagent[26].

The Vis spectra of 22 samples of water including 16 calibration samples marked in grey and 6 test samples marked in blue are presented in the Fig 8S (A). The spectra of calibration sample set are arranged in order of increasing optical density with increasing ammonium nitrogen contencentration. To simulate the optical multisensor systems that are similar to 4-channels OMS applied above, four intervals with a width of 10 or 20 nm were randomly selected from the whole spectra and then avaraged. The interval width was selected based on commercially available LEDs in the visible region. Since the individual LED-channel performances are dramatically different, as it was mentioned in the section on individual regression models for spectrometer and OMS, each channel of the simulated OMSs was multiplied by an arbitrarily selected coefficient from 0 to 1. Fig. 8S (B, C, D, E) shows the resoponses of four simulated OMSs. The intensity ratios at different channels is different as it was observed earlier for OMS-1 and OMS-2 (Fig. 2 B, C).

The PLS-statistics for prediction of ammonium nitrogen in six water samples is presented in Table 4. The PLS-models are constructed using the instrumental responses in 11 samples excluding 5 transfer sets for more convenient comparison with the model transfer results. The original model based on AvaSpec data has a good prediction performance (RMSEP = 0.009 mg/dm^3 with R^2 = 0.99). RMSEP of four simulated OMSs is higher as expected and it is ranged from 0.015 to 0.140 mg/dm^3.

Table 4. PLS statistics for prediction of ammonium nitrogen in six samples of water. The models constructed using the instrumental responses in 11 samples. All PLS models are based on 1 LV.

Equipment	Calibration		Prediction	
	RMSEC	R^2C	RMSEP, mg/dm^3	R^2P
AvaSpec	0.084	0.99	0.009	0.99
Simulated OMS1	0.105	0.99	0.140	0.97
Simulated OMS2	0.119	0.98	0.081	0.99
Simulated OMS3	0.068	0.99	0.015	0.99
Simulated OMS4	0.089	0.99	0.053	0.99

The results of model transfer by DS and SBC are shown in Table 5. Both methods of model transfer work well in these data. For the first two simulated OMSs, the DS works better, while for two others – SBC. The improved accuracy in PLS models of simulated OMSs compared to the original models can be explained by the fact that the model is transferred from the more accurate instrument and the simulation procedure is done in a pretty straightforward way.

Table 5. PLS-results of model transfer from the AvaSpec spectrometer to simulated OMSs. The concentration range of ammonium nitrogen content is 0–3 mg/dm^3. All PLS models are based on 1 LV.

"slave" instrument	Prediction	
	RMSEP, mg/dm^3	R^2P
DS		
Simulated OMS1	0.032	0.99
Simulated OMS2	0.009	0.99
Simulated OMS3	0.028	0.99
Simulated OMS4	0.011	0.99
SBC		
Simulated OMS1	0.109	0.90
Simulated OMS2	0.042	0.99
Simulated OMS3	0.020	0.99
Simulated OMS4	0.007	0.99

The results presented in Tables 4 and 5 confirm real-world applicability of the proposed calibration transfer methodology.

CONCLUSIONS

Obtaining a reliable multivariate calibration model for an optical multisensor system requires significant efforts and investments. Thus, it is highly desirable to have the means of using calibration model derived for one system together with the data acquired with another. While such standardization protocols were already reported for ordinary spectroscopic devices, various optical multisensor systems were not studied in this context so far. We have shown here that multivariate calibration models can be successfully transferred from the full-scale spectrometer to the LED-based OMS or between two OMS. An insignificant loss of accuracy obtained in our study after the transfer of the calibration models compared to the original PLS models proves that DS and SBC methods work quite well, however SBC requires more transfer samples than DS. MLR is more appropriate for model transfer between the OMS, while PLS is better in the case of model transfer from the spectrometer to OMS. The best result has been achieved for MLR model using DS for transfer between OMSs (RMSEP=0.004). These results are important for further development of optical multisensor system especially in the context of industrial and express analysis of various media.

This study confirms the feasibility of calibration model transfer between analytical instruments with dramatically different numbers of variables using simple approaches and small datasets.

ASSOCIATED CONTENT

Supporting Information

Supplementary Table 1S. The reproducibility of the OMS-1 for 3 samples of water.

Supplementary Table 2S. Composition of standard sets (M) prepared by diagonal design.

Supplementary Table 3S. PLS regression results on full dataset (20 samples) for cobalt(II) nitrate.

Supplementary Table 4S. Results of model transfer from OMS to spectrometer (original models for Q-mini and OMS are shown in Table 1).

Supplementary Figure 1S. Spectra of four LEDs from OMS-1 acquired on Q-mini.

Supplementary Figure 2S. The reproducibility of OMS-1 for (A) channel 280 nm, (B) channel 340 nm, (C) channel 600 nm, and (D) channel 280 nm.

Supplementary Figure 3S. Diagonal design scheme with sample IDs in circles; samples for model transfer are marked in black.

Supplementary Figure 4S. The cross-correlation with Q-mini spectrometer at the LEDs wavelengths for OMS-1.

Supplementary Figure 5S. The cross-correlation with Q-mini spectrometer at the LEDs wavelengths for OMS-2.

Supplementary Figure 6S. The cross-correlation between OMS-1 and OMS-2.

Supplementary Figure 7S. PLS-model transfer between OMS-1 and OMS-2 using SBC: Predicted versus measured values for copper model, where values from ‘‘master” instrument (OMS-1) are designated by blue triangles, wrong and corrected values from ‘‘slave” instrument (OMS-2) are designated by red circles and green squares, respectively.

Supplementary Figure 8S. The measurement results from 22 water samples (with six test samples marked blue) obtained with: (A) AvaSpec spectrometer (in the range 365–425 nm), (B) simulated OMS1, (C) simulated OMS2; (D) simulated OMS3, and (E) simulated OMS4.

AUTHOR INFORMATION

Corresponding Author

* Anastasiia Surkova, Email: melenteva-anastasija@rambler.ru, ORCID: 0000-0001-6220-7822

Author Contributions

The manuscript was written through contributions of all authors.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGEMENTS

A. Surkova would like to acknowledge the financial support from the Russian Science Foundation, grant #19-79-00076 for the calculation of the model transfer. D. Kirsanov and A. Legin acknowledge the support of their work on optical multisensor system by the Russian Science Foundation grant #18-19-00151. D. Kirsanov also acknowledges DAAD Dmitrij Mendeleev-Programm, 2018 (57402705) for supporting

his research stay in Aalen, Germany. A. Bogomolov thanks the Ministry of Education and Science of the Russian Federation within the framework of state task No. 0778-2020-0005 for planning the experiment.

REFERENCES

(1) Valle, M. Electronic tongues employing electrochemical sensors. *Electroanalysis* **2010**, *22*, 1539–1555. https://doi:10.1002/elan.201000013.

(2) Campos, I.; Alcañiz, M.; Aguado, D.; Barat, R.; Ferrer, J.; Gil, L.; Marrakchi, M.; Martínez-Mañez, R.; Soto, J.; Vivancos, J.-L. A voltammetric electronic tongue as tool for water quality monitoring in wastewater treatment plants. *Water Res.* **2012**, *46* (8), 2605–2614. https://doi.org/10.1016/j.watres.2012.02.029.

(3) Ul Alam, A.; Clyne, D.; Jin, H.; Hu, N.-X.; Deen M. J. Fully integrated, simple and low-cost electrochemical sensors array for in situ water quality monitoring. *ACS Sens.* **2020**, *5* (2), 412–422. https://doi.org/10.1021/acssensors.9b02095.

(4) Askim, J. R.; Mahmoudi, M.; Suslick, K. S. Optical sensor arrays for chemical sensing: the optoelectronic nose. *Chem. Soc. Rev.* **2013**, *42* (22), 8649–8682. https://doi.org/10.1039/c3cs60179j.

(5) Tortora, L.; Stefanelli, M.; Mastroianni, M.; Lvova, L.; Di Natale, C.; D'Amico, A.; Filippini, D.; Lundstrom, I.; Paolesse, R. The hyphenated CSPT-potentiometric analytical system: An application for vegetable oil quality control. *Sens. Actuators B: Chem.* **2009**, *142* (2), 457–463. https://doi.org/10.1016/j.snb.2009.05.022.

(6) Paolesse, R.; Alimelli, A.; D'Amico, A.; Venanzi, M.; Battistini, G.; Montalti, M.; Filippini, D.; Lundström, I.; Di Natale, C. Insights on the chemistry of a,c-biladienes from a CSPT investigation. *New J. Chem.* **2008**, *32* (7), 1162–1166. https://doi.org/10.1039/B800512E.

(7) Rukosueva, E. A.; Dobrolyubov, E. O.; Goryacheva, I. Y.; Beklemishev, M. K. Discrimination of whiskies using an "Add-a-Fluorophore" fluorescent fingerprinting strategy. *Microchem. J.* **2019**, *145*, 397–405. https://doi.org/10.1016/j.microc.2018.11.002.

(8) Cole, M.; Spulber, I.; Gardner, J. W. Surface acoustic wave electronic tongue for robust analysis of sensory components. *Sens. Actuator A: Phys.* **2015**, *207*, 1147–1153. https://doi.org/10.1016/j.snb.2014.09.029

(9) Chung, J.; Chen, Y.; Kim, S.-J. High-density impedance-sensing array on complementary metal-oxide-semiconductor circuitry assisted by negative

dielectrophoresis for single-cell-resolution measurement. *Sens. Actuators B: Chem.* **2018**, *266*, 106–114. https://doi.org/10.1016/j.snb.2018.03.113

(10) Bogomolov, A.; Ageev, V.; Zabarylo, U.; Usenov, I.; Schulte, F.; Kirsanov, D.; Belikova, V.; Minet, O.; Feliksberger, E.; Meshkovsky, I.; Artyushenko, V. LED-based near infrared sensor for cancer diagnostics. *Proc. SPIE* **2016**, *9715*, 971510. https://doi.org/10.1117/12.2214342.

(11) Giovenzana, V.; Civelli, R.; Beghi, R.; Oberti, R.; Guidetti, R. Testing of a simplified LED based vis/NIR system for rapid ripeness evaluation of white grape (*Vitis vinifera L.*) for Franciacorta wine. *Talanta* **2015**, *144*, 584–591. https://doi.org/10.1016/j.talanta.2015.06.055.

(12) Dong, Y.; Liu, X.; Mei, L.; Feng, C.; Yan, C.; He, S. LED-induced fluorescence system for tea classification and quality assessment. *J. Food Eng.* **2014**, *137*, 95–100. https://doi.org/10.1016/j.jfoodeng.2014.03.027.

(13) Sjöström, M.; Wold, S.; Lindberg, W.; Martens, H. A multivariate calibration problem in analytical chemistry solved by partial least squares models in latent variables. *Anal. Chim. Acta* **1983**, *150*, 61–70. https://doi.org/10.1016/S0003-2670(00)85460-4.

(14) Bogomolov, A. Diagonal designs for a multi-component calibration experiment, Anal. Chim. Acta. **2017**, *951,* 46–57. https://doi.org/10.1016/j.aca.2016.11.038.

(15) Kirsanov, D.; Panchuk, V.; Agafonova-Moroz, M.; Khaydukova, M.; Lumpov, A.; Semenov, V.; Legin, A. A sample-effective calibration design for multiple components. *Analyst* **2014,** *139* (17), 4303–4309. https://doi.org/10.1039/c4an00227j.

(16) Melenteva, A.; Galyanin, V.; Savenkova, E.; Bogomolov, A. Building global models for fat and total protein content in raw milk based on historical spectroscopic data in the visible and short-wave near infrared range. *Food Chem.* **2016**, *203*, 190–198. https://doi.org/10.1016/j.foodchem.2016.01.127.

(17) Wang, L.-J.; Chang, Y.-C.; Ge, X.; Osmanson, A. T.; Du, D.; Lin, Y.; Li, L. Smartphone optosensing platform using a DVD grating to detect neurotoxins. *ACS Sens*. **2016**, *1* (4), 366–373. https://doi.org/10.1021/acssensors.5b00204.

(18) Bouveresse, E.; Massart, D. L. Standardisation of near-infrared spectrometric instruments : A review. *Vib. Spectrosc.* **1996**, *11*, 3–15. https://doi.org/10.1016/0924-2031(95)00055-0.

(19) Feudale, R. N.; Woody, N. A.; Tan, H.; Myles, A. J.; Brown, S. D.; Ferré, J. Transfer of multivariate calibration models: a review. *Chemom. Intell. Lab. Syst.* **2002**, *64* (2), 181–192. https://doi.org/10.1016/S0169-7439(02)00085-0.

(20) Wang, Y. D.; Veltkamp, D. J.; Kowalski, B. R. Multivariate instrument standardization. *Anal. Chem.* **1991**, *63* (23), 2750–2756. https://doi.org/10.1021/ac00023a016.

(21) Wang, Y.; Kowalski, B. R. Calibration transfer and measurement stability of near-infrared spectrometers. *Appl. Spectrosc.* **1992**, *46*, 764–771. https://doi.org/10.1366/0003702924124808.

(22) Panchuk, V.; Kirsanov, D.; Legin, A.; Oleneva, E.; Semenov, V. Calibration transfer between different analytical methods. *Talanta* **2017**, 170, 457–463. https://doi:10.1016/j.talanta.2017.04.039.

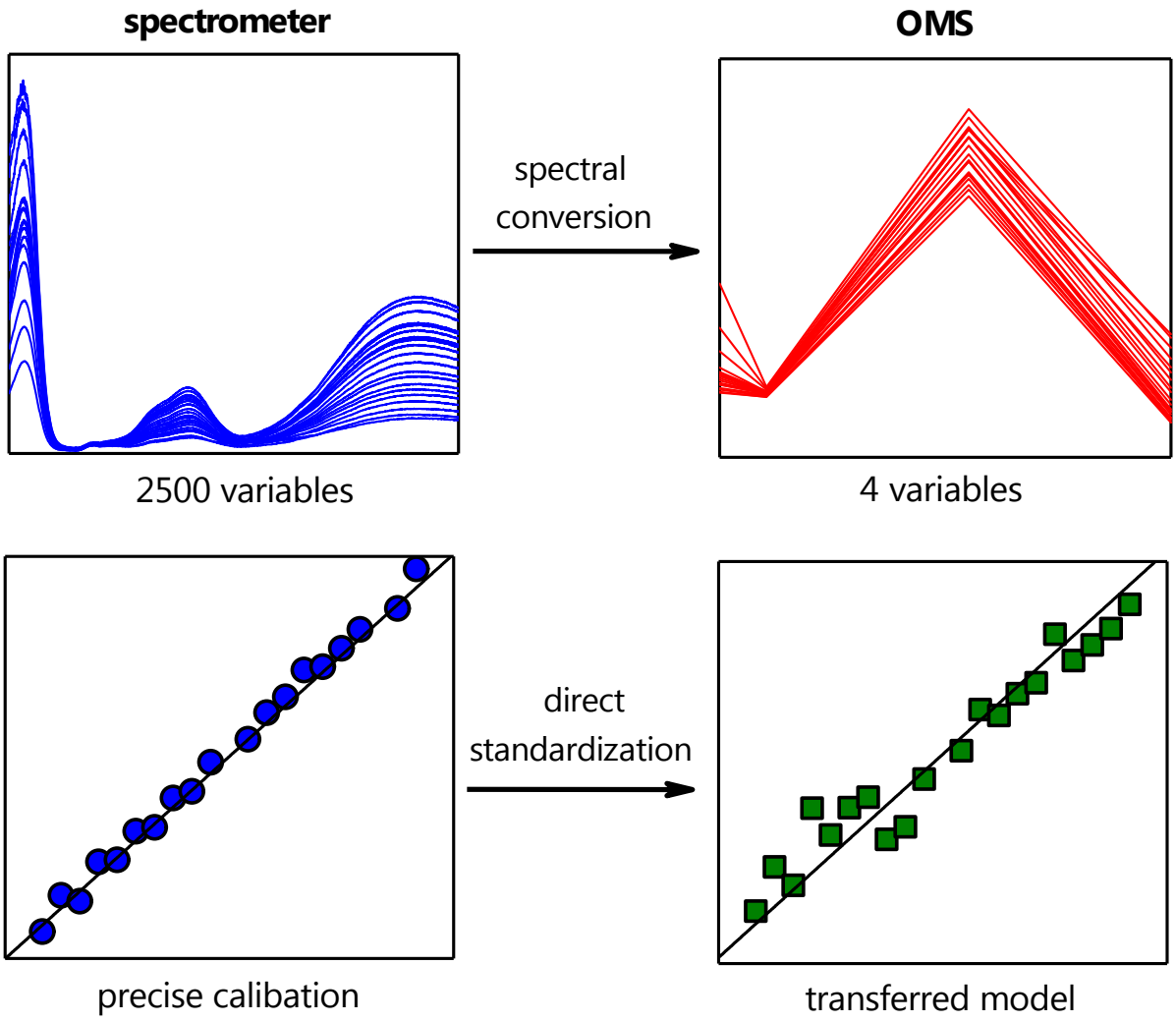
(23) Khaydukova, M.; Medina-plaza, C.; Rodriguez-mendez, M. L.; Panchuk, V.; Kirsanov, D.; Legin, A. Multivariate calibration transfer between two different types of multisensor systems. *Sens. Actuators B: Chem.* **2017**, 246, 994–1000. https://doi:10.1016/j.snb.2017.02.099.

(24) Esbensen, K. H. Multivariate data analysis – in practice (5th ed.). CAMO Process AS, Oslo, 2001.

(25) Wold, S.; Sjöström, M.; Eriksson, W. PLS-regression: a basic tool of chemometrics, *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. https://doi.org/10.1016/S0169-7439(01)00155-1.

(26) GOST 33045-2014. Water. Methods for determination of nitrogen-containing matters, **2019**.

for TOC only


spectrometer
OMS
spectral
conversion
2500 variables
4 variables
direct
standardization
precise calibation
transferred model