

Analyst

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.



Analyst

ARTICLE

Calibration transfer via extreme learning machine auto-encoder

Wo-Ruo Chen^a, Jun Bin^b, Hong-Mei Lu^a, Zhi-Min Zhang^{a,*}, Yi-Zeng Liang^{a,†}

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

In order to solve the spectra standardization problem in near-infrared (NIR) spectroscopy, Transfer via Extreme learning machine Auto-encoder Method (TEAM) has been proposed in this study. A comparative study among TEAM, piecewise direct standardization (PDS), generalized least squares (GLS) and calibration transfer method based on canonical correlation analysis (CCA) was conducted, and the performances of these algorithms were benchmarked with three spectral dataset: corn, tobacco and pharmaceutical tablets spectra. Results show that TEAM is a stable method and can significantly reduce prediction errors compared with PDS, GLS and CCA. TEAM can also achieve the best RMSEPs in most cases with small number of calibration set. TEAM is implemented in Python language and available as open source package at <https://github.com/zmzhang/TEAM>.

Introduction

NIR spectroscopy has become a popular analytical technique in many industrial applications, like petrochemical, agricultural, pharmaceutical and etc.¹⁻⁴. The NIR is broadly used in online process monitoring applications. However, the NIR spectra may contain instrument-related variation not captured by the original model and this can lead to erroneous predictions. In general, the model obtained from one spectrometer is not applicable to other spectrometers. Recalibration can be used to solve this problem. But the recalibration process would be both costly and time consuming. A more acceptable way is to do calibration transfer, which can correct the difference of spectra between the master instrument and another (slave) instrument. Essentially, spectra on the slave instruments are transformed so as to appear as if originating from the master instrument; then the original calibration model can be used on the transformed spectra.

Different calibration transfer techniques have been developed over the past years, which can be divided into three main approaches⁵. The first is model updating that is to rebuild the model with the addition of a few samples measured under new conditions to the old calibration set measured on primary conditions. But the samples added must contain the variability in the new conditions such that new model spans both old and new experimental conditions⁶. The second kinds of calibration transfer method is to reduce the difference of data measured on

different conditions by signal preprocessing methods, including baseline elimination, derivative techniques, multiplicative scatter correction (MSC)⁷, finite impulse response (FIR) filtering⁸, orthogonal signal correction (OSC)⁹, and generalized least squares (GLS)¹⁰.

The third kind of method is standardization methods. Through simple univariate slope and bias correction (SBC)¹¹, the predict values can be standardized. It's based on the assumption that the predictions values of slave instruments and master instruments are linear dependence. Then the predicted values for the new samples can be corrected for the bias and the slope of the regression equation. Alternatively, the spectra from the slave instrument can be corrected to become closer to the spectra of the same standardization samples from the master instrument, and the original model can be used for prediction on the slave instrument and new calibration is not required, e.g. direct standardization (DS), piecewise direct standardization (PDS) method developed by Wang et al¹², as well as the patented method developed by Shenk and Westerhaus¹³. The third approach is to try to standardize of the model coefficients that was also proposed by Wang et al.¹² or by Forina et al¹⁴.

Among all the methods mentioned above, PDS is one of the most widely used transfer methods, and is typically employed as a reference for other novel techniques. Its superiority over other standardization methods can be attributed to its local character and multivariate nature⁵, enabling simultaneous correction of intensity differences, wavelength shifts and peak broadening. However, PDS method employs PCR or PLS regression method to transform the whole spectra also including useful parts which make contribution to the model and unuseful parts.

Recently, a model transfer method based on canonical correlation analysis¹⁵ (CCA) was proposed. CCA was successfully applied to correct the differences between spectra measured on different instruments because of its ability to reveal the correlations between them. However, the

^a College of Chemistry and Chemical Engineering, Central South University, Changsha 410083, China.

^b College of Bioscience and Biotechnology, Hunan Agriculture University, Changsha 410128, China.

* To whom correspondence should be addressed, Email: zmzhang@csu.edu.cn.

† Correspondence author, Yi-Zeng Liang, Email: yizeng.liang@263.net.

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

performance of CCA is sometime fluctuate and may face the problem that it may fail when the spectral matrix is close to singular. It is worth noting that both CCA and PDS methods use linear method to find the relationship between the slave spectra and master spectra. Besides the differences can be fitted by local PCR, local PLS and CCA, there may exist nonlinear differences between slave and master instruments which PDS and CCA cannot fit.

As existing of intensity differences, wavelength shifts and peak broadening between the slave spectra and master spectra, we suppose that the spectra of slave instruments and master instruments is nonlinear dependence. In order to find the relationship between them, a novel calibration transfer method has been proposed based on ensemble ELM (TEAM) in this study. Because of the universally approximate ability of ELM¹⁶⁻¹⁸, the relationship between slave spectra and master spectra can be found by TEAM. Extreme learning machine is a new learning scheme of feedforward neural networks proposed by Huang et al^{16, 19}. Different from traditional feedforward networks algorithms which use the slow gradient based learning algorithms and tune the parameters of the networks iteratively, ELM randomly chooses the input weights and analytically determines the output weights to provide the best generalization performance at extremely high speed. ELM with arbitrary assigned input weights and hidden layer biases and with almost any nonzero activation function can universally approximate any continuous functions on any compact input sets²⁰.

ELM has both universal approximation and classification capabilities with significant advantages such as fast learning speed, ease of implementation and minimal human intervention. Compared with PDS algorithm in based on the local linear regression, TEAM can be used to establish the relationship between slave spectra and master spectra, which may significantly reduce prediction errors after calibration transfer. Due to advantages of TEAM, it can be used to correct the systematic difference between signals produced by master instruments and slave instruments. It can be applied to various analytical techniques, and TEAM has been used to transfer NIR dataset to illustrate its performance and advantages in this study.

Theory

Notation

For each spectral matrix $\mathbf{X}_{N \times M}$ ($\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$), the rows N and columns M of \mathbf{X} correspond to the samples and spectral variables respectively. $\mathbf{X}_{(i,j)}$ represents the i^{th} sample at j^{th} wavelength. A superscript letter is added in front of the matrix to distinguish the spectra from the master instrument or the slave instrument, that is, ${}^m\mathbf{X}$ and ${}^s\mathbf{X}$ denote the spectra from master instrument and slave instrument respectively. Moreover, ${}^m\mathbf{C}$ (${}^m\mathbf{V}$, ${}^m\mathbf{P}$) and ${}^s\mathbf{C}$ (${}^s\mathbf{V}$, ${}^s\mathbf{P}$) correspond to calibration (validation, independent test) samples of different instruments.

Extreme learning machine

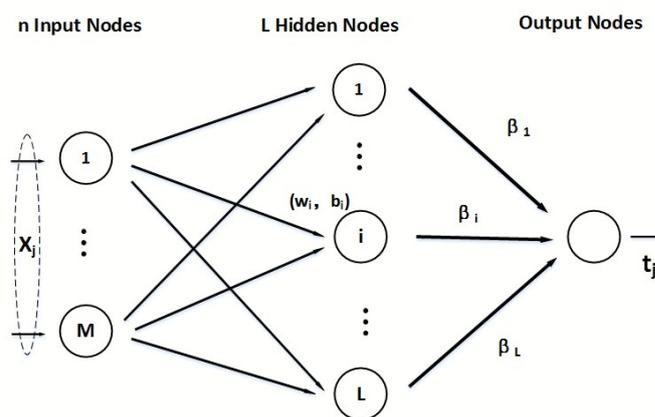


Fig. 1 the network structure of ELM

For Single-Hidden Layer Feedforward Networks (SLFNs), the ELM algorithm can provide efficient unified solutions. Unlike the back-propagation or conjugate gradient descent training algorithm, the theory of ELM shows that the hidden nodes of generalized feedforward networks are important but needn't be tuned and these hidden nodes can be randomly generated. All the hidden node parameters are independent from the target functions or the training datasets. The input data is mapped to L -dimensional ELM random feature space. Fig.1 shows the network structure of ELM. ELM algorithm can be summarized as Fig.1.

Algorithm ELM: Given activation function $g(x)$, a training set $N = \{(\mathbf{x}_j, \mathbf{t}_j) | \mathbf{x}_j \in \mathbf{R}^n, \mathbf{t}_j \in \mathbf{R}^m, j = 1, \dots, N\}$, and hidden node number L .
Step 1: Randomly assign input weight \mathbf{w}_i and bias b_i ($i = 1, 2, \dots, L$).
Step 2: Calculate the hidden layer output matrix \mathbf{H} .

$$\sum_{i=1}^L \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{t}_j, j = 1, \dots, N \quad (1)$$

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{T} \quad (2)$$

Where

$$\mathbf{H}(\mathbf{w}_1, \dots, \mathbf{w}_L, b_1, \dots, b_L, \mathbf{x}_1, \dots, \mathbf{x}_N) = \begin{pmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \dots & g(\mathbf{w}_L \cdot \mathbf{x}_1 + b_L) \\ \vdots & \dots & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \dots & g(\mathbf{w}_L \cdot \mathbf{x}_N + b_L) \end{pmatrix}_{N \times L} \quad (3)$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{pmatrix}_{L \times M} \quad \text{and} \quad \mathbf{T} = \begin{pmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_N^T \end{pmatrix}_{N \times M}$$

Step 3: Calculate the output weight by, here is the Moore-Penrose generalized inverse of \mathbf{H} and $\mathbf{T} = (\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_N)^T$.

$$\boldsymbol{\beta} = \mathbf{H}^+ \mathbf{T} \quad (4)$$

Analyst

To improve generalization performance and make the solution more robust, we can add a regularization term²¹ as shown:

$$\beta = \left(\frac{\mathbf{I}}{C} + \mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{T} \quad (5)$$

Where C is regularization parameter to make the result solution more stabler and have better generalization performance¹⁶.

The ELM algorithm is an extremely fast algorithm which has both universal approximation and classification capabilities.

ELM auto-encoder

The ELM based auto-encoder (ELM-AE) can be seen as a special case of ELM, where the input data is used as output data ($\mathbf{t}=\mathbf{x}$), and the randomly generated weights and biases of the hidden nodes are chosen to be orthogonal. Orthogonalization of these randomly generated hidden parameters tends to improve ELM-AE's generalization performance. According to ELM theory, ELMs are universal approximators, hence ELM-AE is as well. Fig.2 shows ELM-AE's network structure. In ELM-AE, the orthogonal random weights and biases of the hidden nodes project the input data to a different or equal dimension space, as shown by the Johnson-Lindenstrauss lemma²² and calculated as

$$\begin{aligned} \mathbf{h} &= g(\mathbf{w} \cdot {}^s \mathbf{x} + \mathbf{b}) \\ \mathbf{w}^T \mathbf{w} &= \mathbf{I}, \mathbf{b}^T \mathbf{b} = 1 \end{aligned} \quad (6)$$

Where $\mathbf{w}=[w_1, \dots, w_L]$ are the orthogonal random weights, and $\mathbf{b}=[b_1, \dots, b_L]$ are the orthogonal random biases between the input and hidden nodes. ELM-AE's output weight β is responsible for learning the transformation from the feature space to input data. For sparse and compressed ELM-AE representations, we calculate output weights β as follows:

$$\beta = \left(\frac{\mathbf{I}}{C} + \mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{H}^T {}^m \mathbf{X} \quad (7)$$

Where $\mathbf{H}=[h_1, \dots, h_N]$ are ELM-AE's hidden layer outputs, and $\mathbf{X}=[x_1, \dots, x_N]$ are its input and output data, C is regularization parameter.

In our study, the master and slave spectra are acquired from the same samples but on different spectrometers (slave and master instruments). The master spectra and slave spectra have essentially identical chemical information; their differences are mainly originated from the systematic differences between instruments. Therefore, ELM-AE has been used to find the transfer relationship between these instruments and the spectra on slave instrument is chosen as input and the spectra on master instrument as output.

Calibration transfer by Ensemble ELM-AEs

With the growth of hidden nodes, the prediction performance is getting better but the result is fluctuating in small scale. So

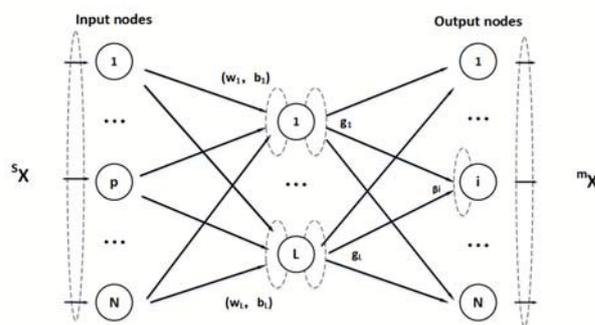


Fig.2-the network structure of ELM-AE for calibration transfer

randomly initialized hidden nodes for multiple times, we train multiple ELMs on the same training dataset to get a series of spectra. By averaging these spectra, the final transfer spectrum can be obtained. This procedure is usually known as ensemble and this network is called as Ensemble Extreme Learning Machines (EELM). This procedure improves the stability of the transfer result. Overview of TEAM is given in Fig.3.

Step1: the samples are divided into three parts: calibration set, validation set and independent test set ${}^m \mathbf{C}({}^m \mathbf{V}, {}^m \mathbf{P})$. Then, the same subset samples are collected on the secondary instrument ${}^s \mathbf{C}({}^s \mathbf{V}, {}^s \mathbf{P})$.

Step 2: randomly select hidden nodes number, for example from 300 to 500(more details about how to optimize the number of hidden neurons and how to choose a correct interval can be seen in supplement information). Using validation set to optimize the value of regularization parameter C . Training multiple ELM-AEs on ${}^m \mathbf{C}$ and ${}^s \mathbf{C}$. For each ELM-AE, the relationship between ${}^m \mathbf{C}$ and ${}^s \mathbf{C}$ are settled with following equations:

$$\sum_{i=1}^L \beta_i g(\mathbf{w}_i \cdot {}^s \mathbf{C}(j, :) + b_i) = {}^m \mathbf{C}(j, :), j=1, \dots, M \quad (8)$$

Step 3: Then for each ELM-AE model trained on ${}^m \mathbf{C}$ and ${}^s \mathbf{C}$ transfer spectra ${}^{new} \mathbf{P}_k$ ($k=1, 2, \dots, K$) can be obtained, K is the trained ELMs numbers:

$$\sum_{i=1}^L \beta_i g(\mathbf{w}_i \cdot {}^s \mathbf{P}(j, :) + b_i) = {}^{new} \mathbf{P}_k(j, :), j=1, \dots, M \quad (9)$$

Step 4: Get the average transfer spectra ${}^{average} \mathbf{P}$ from a series of transfer spectra, ${}^{average} \mathbf{P}$ is the transfer spectra of independent test set from the slave instrument to master instrument.

$${}^{average} \mathbf{P} = \frac{\sum_{k=1}^K {}^{new} \mathbf{P}_k}{m} \quad (10)$$

Experimental

Spectral dataset

In this study, three NIR datasets were used to investigate the performance of our method TEAM, PDS, GLS as well as CCA method.

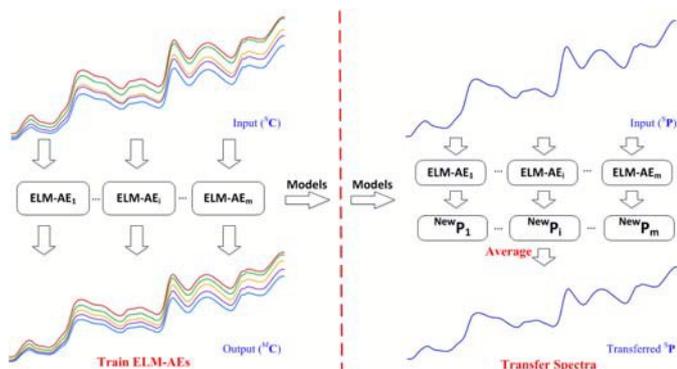


Fig.3-The system diagram of TEAM. ^SC and ^MC are the spectra of master and slave instruments respectively. ^SC and ^MC are sent to ELM-AEs to generate ELM-AE models. ^SP are spectra of slave instruments which is needed to be transferred by the trained ELM-AEs. $^{\text{New}}\text{P}_i$ are the transferred spectra by the i th ELM-AE. The final transferred spectra of ^SP can be obtained by averaging all the $^{\text{New}}\text{P}_k$.

The first dataset contains spectra of 80 corn samples measured on different NIR spectrometers (MP5 for 'primary'; M5 and MP6 for 'secondary'). The wavelength ranges from 1100 to 2498nm at 2nm intervals (700 channels). The moisture value for each sample is also included. MP5spec is the spectral measured on primary instrument FOSS NIRsystems 5000. M5spec, MP6spec are the spectra measured on slave instruments FOSS NIRsystems 5000 and FOSS NIRsystems 6000, respectively. The data are available at <http://www.eigenvector.com/Data/Corn/>.

The second dataset consists of 258 tobacco samples measured on two different NIR instruments. The Primary spectral data were measured on the AntarisTM Fourier transform (FT) instrument (Nicolet Inc., USA). The reflectance range from 10000 to 4000 cm^{-1} at 4 cm^{-1} interval (1557 channels) and the number of scans per spectrum was set to 32. The secondary spectral dataset was measured on the SpectrumTM Fourier transform (FT) instrument (PerkinElmer Inc., USA). The spectra were also acquired in the reflectance mode from 10000 to 4000 cm^{-1} at 4 cm^{-1} interval (1557 channels) and the number of scans per spectrum was set to 16. The concentration of total nitrogen was measured on a Skalar SAN plus segmented flow analyzer (Skalar Analytical Instruments, Netherlands).

The third data set is a public available dataset for calibration transfer from the IDRC shootout 2002. Spectra of 654 pharmaceutical tablets from two spectrometers (Foss NIRsystems, Silver-spring, MD) are measured in the transmittance mode. Tablets data from two instruments have been split into two calibration sets (155 tablets, Calibrate 1 and Calibrate 2) and two test sets (460 tablets, Test 1 and Test 2). The dataset includes tablets with a wide ASSAY range, 152-239mg, for developing calibration model. The spectral region of interest for calibration transfer for the ASSAY content is 1100-1700nm.

Implementation and Data processing

All the chemometric methods used for processing data were implemented by our research group in Python programming language (version 2.7.4).

Calibrations were performed by partial least squares (PLS) regression. 10-fold cross validation was used to determine the optimal number of latent factors. Prediction performance was evaluated by a root mean square error on independent test set (RMSEP). The samples was split into three sets: calibration set, validation set and independent test set. The validation set is used to evaluate which activation function is better and to optimize the value of regularization parameter C. The independent test set is used to assess the performance of calibration transfer method. First the samples was split into two set by Kennard-Stone²³ split method, one is train set and the other is independent test set. Then the train set was split into two set by Kennard-Stone split method, one is calibration set and the other is validation set. For corn, 80 samples was split into three sets: 48 samples for calibration samples, 16 samples for validation samples and 16 samples for independent test samples. For tobacco, 258 samples were split into three sets: 154 samples for calibration samples, 52 samples for validation samples and 52 samples for independent test samples. For pharmaceutical tablets, samples have already been partitioned; we mixed it and then split into three sets: 305 samples for calibration samples, 155 samples for validation samples and 155 samples for independent test samples. The PLS model was set on the calibration samples. The Kennard-Stone method was adopted to select standardization samples on calibration samples.

Results and Discussion

To evaluate the performance of TEAM, the commonly used methods CCA, GLS and PDS are tested using the three data sets along with recalibration on slave instrument for comparison. For all of three datasets, the parameters of TEAM are set as follows: input weight w_i and bias b_i are set randomly range from -0.1 to 0.1, the hidden nodes is selected 50 elements from 300 to 500, the activation function is chosen as tanh, the regularization coefficient is optimized by validation set and the default value is set as 50000. For each hidden node, ELM-AE model is trained to transfer the spectra on slave towards the master. Then, the average transfer spectra can be obtained. Finally, the calibration models are used to predict the concentration of interest and calculate the RMSEP for evaluating the performance of transferring.

Performance Comparison

Corn Dataset

The Fig.S-2 and Fig.S-4 in supporting information show the differences between the 16 predictive corn samples of SP and MP by different calibration transfer methods, respectively. From these plots, one can see that the differences between the transfer spectra and the master spectra are significantly smaller than the ones between the slave spectra and the master spectra of the independent test sets. It is worth noticing that TEAM exhibits the smallest differences between the spectra. The performances of the four methods on corn samples are shown in Table 1, Fig.S-3 and Fig.S-5 in the supporting information. The

results listed in Table 1 show clearly that both TEAM and CCA have much lower prediction errors than PDS and GLS. In the table, w stands for the window size of PDS method and N stands for the number of subset samples. To evaluate the effect of the number of subset samples on different calibration methods, 10, 15, 20, 25, 30, 35, 40 standardization samples are considered. From the table we can see that different sizes of windows are considered for PDS method. From the results, one

can infer that the small window size ($w=1$) gives a better result from M5 to MP5, but for MP6 to MP5, we can see that the large window size ($w=7$) gives the better results. With the increasing of subset samples, the performance of TEAM is better, but the change trend of the performance of CCA, PDS and GLS is not so clear. From Fig.S-3 and Fig.S-5 and Table 1, one can see clearly that TEAM gets the best performance for prediction than other three methods.

Table.1 RMSEP results of three spectra dataset with different transfer methods

RMSEP for corn data (M5to MP5)									
RMSEP	PDS				CCA	GLS	TEAM	PLS_M	Recalibration
	W=1	W=3	W=5	W=7					
N=10	0.412	0.412	0.414	0.417	0.131	0.199	0.186		
N=15	0.410	0.410	0.412	0.414	0.161	0.229	0.098		
N=20	0.442	0.443	0.445	0.450	0.148	0.202	0.098		
N=25	0.434	0.435	0.437	0.440	0.192	0.194	0.121	0.130	0.007
N=30	0.422	0.423	0.425	0.428	0.117	0.234	0.083		
N=35	0.412	0.413	0.415	0.417	0.140	0.231	0.078		
N=40	0.415	0.416	0.418	0.421	0.084	0.230	0.073		
RMSEP for corn data (MP6 to MP5)									
N=10	0.584	0.566	0.550	0.535	0.170	0.177	0.236		
N=15	0.575	0.559	0.544	0.523	0.146	0.167	0.141		
N=20	0.604	0.584	0.567	0.550	0.194	0.170	0.158		
N=25	0.593	0.574	0.558	0.543	0.168	0.169	0.146	0.130	0.133
N=30	0.589	0.572	0.556	0.541	0.151	0.172	0.146		
N=35	0.586	0.570	0.554	0.539	0.133	0.170	0.140		
N=40	0.587	0.571	0.555	0.540	0.152	0.166	0.136		
RMSEP for tobacco data									
N=10	0.163	0.163	0.162	0.162	0.160	0.071	0.141		
N=15	0.158	0.158	0.158	0.157	0.096	0.072	0.084		
N=20	0.169	0.169	0.169	0.168	0.081	0.074	0.073		
N=25	0.170	0.170	0.169	0.169	0.073	0.073	0.073		
N=30	0.176	0.176	0.176	0.175	0.067	0.072	0.072		
N=35	0.168	0.168	0.167	0.167	0.066	0.072	0.073	0.072	0.064
N=40	0.168	0.168	0.167	0.167	0.063	0.073	0.069		
N=50	0.162	0.162	0.161	0.161	0.065	0.067	0.068		
RMSEP for pharmaceutical tablet data									
N=10	6.82	7.13	7.03	7.08	5.18	5.17	5.48		
N=15	6.23	6.49	6.42	6.48	Nan*	4.71	6.94		
N=20	6.22	6.49	6.41	6.46	Nan	4.77	5.24		
N=25	6.19	6.43	6.35	6.39	Nan	4.58	4.52		
N=30	6.53	6.80	6.71	6.73	Nan	4.51	4.37		
N=35	6.58	6.87	6.77	6.77	Nan	4.52	4.34	4.23	4.57
N=40	6.32	6.59	6.49	6.51	Nan	4.59	4.34		
N=50	6.59	6.81	6.72	6.74	Nan	4.89	4.34		
N=55	6.47	6.68	6.60	6.62	4.62	4.88	4.34		
N=60	6.36	6.56	6.48	6.50	4.68	4.67	4.32		

* As the matrix is nearly to singular, so the CCA cannot transfer the spectra of slave instrument to master instrument

Table.2 RMSEP for three spectra dataset of TEAM with different activation functions

RMSEP	TEAM							
	corn data (M5 to MP5)		corn data (MP6 to MP5)		tobacco data		Pharmaceutical tablet data	
	sigmoid	tanh	sigmoid	tanh	Sigmoid	tanh	sigmoid	tanh
N=10	0.120	0.182	0.208	0.250	0.114	0.129	6.52	6.18
N=15	0.123	0.117	0.173	0.164	0.076	0.077	7.22	7.10
N=20	0.100	0.077	0.163	0.145	0.064	0.068	6.28	6.09
N=25	0.100	0.090	0.160	0.154	0.062	0.063	5.82	5.63
N=30	0.077	0.049	0.154	0.160	0.062	0.064	5.67	5.61
N=35	0.058	0.063	0.159	0.164	0.062	0.064	5.63	5.57
N=40	0.072	0.102	0.157	0.158	0.064	0.065	5.64	5.57
N=45	0.067	0.092	0.157	0.147	0.064	0.065	5.62	5.53

Table.3 RMSEP for three spectra dataset of TEAM with different regularization parameters

RMSEP	TEAM							
	corn data (M5 to MP5)		corn data (MP6 to MP5)		tobacco data		Pharmaceutical tablet data	
	0	50000	0	50000	0	50000	500	50000
N=10	0.238	0.182	0.266	0.250	0.137	0.129	6.18	5.92
N=15	0.220	0.117	0.160	0.164	0.089	0.077	7.10	6.85
N=20	0.124	0.077	0.168	0.145	0.077	0.068	6.09	5.84
N=25	0.195	0.090	0.190	0.154	0.066	0.063	5.63	5.62
N=30	0.177	0.049	0.167	0.160	0.061	0.064	5.61	5.66
N=35	0.169	0.063	0.180	0.164	0.063	0.064	5.57	5.62
N=40	0.183	0.102	0.176	0.158	0.064	0.065	5.57	5.60
N=45	0.135	0.092	0.161	0.147	0.063	0.065	5.53	5.57

Tobacco Dataset

The results of this dataset can be seen from Fig. S-7 to Fig. S-8 in the supporting information as well as Table 1. With the increasing of subset samples, the change trend of performance of PDS is not so clear, but the performance of TEAM is better. And with the change of window sizes, the result of PDS cannot get better.

When the number of subset samples is larger than 20, the performance of TEAM is stable and close to the RMSEP result of independent test set on master instrument. CCA, GLS and TEAM methods all can get good performance on tobacco dataset.

Pharmaceutical tablets dataset

From Fig. S-10 to Fig. S-11 in supporting information and also table 1 one can also see that the performance of TEAM is the best and the performance of GLS is rather better than PDS. When the number of subset samples is larger than 30, the performance of TEAM is stable. It is worth noting that CCA cannot well transfer spectra with the numbers of subset samples are 15, 20, 25, 30, 35, 40 and 50, respectively, as the matrices are singular for this dataset (see Table 1). When the number of subset samples is larger than 35, the performance of TEAM is stable and close to the RMSEP result of independent test set on master instrument.

From the discussion above, one can easily reach that TEAM can achieve the best RMSEPs in most cases with small number

of calibration set for establishing the transfer relationship. Therefore, the performance of TEAM is the best as it can find more accurate transfer relationship between the slave spectra and master spectra with small number of spectra in calibration set to reduce the effect from systematic differences between instruments.

Stability of ensemble ELM-AEs

In order to check the stability of the proposed method, the weights and biases of the hidden nodes were generated randomly. The number of hidden nodes for each ELM-AE is also randomly selected from 300 to 500 for 10 times. 2000 ensemble ELM-AEs with 40 standardization samples were trained for tobacco dataset, and then transferred the independent test set with these ELM-AEs to obtain the distribution of the RMSEPs. Fig.4 shows the distribution of the RMSEPs. T-test was performed and at the significance level $\alpha = 0.05$, the mean of RMSEP obtained with TEAM is significantly less than the RMSEP obtained with PDS and GLS. The variance of the distribution is 0.0025, which is relatively small. One can infer from Fig.4 that TEAM is a stable calibration transfer method even the weights, bias and the number of hidden nodes are randomly generated. The same results can be obtained by other datasets. Please see Fig S-12,

Fig S-13 and Fig S-14 in the supplement information for more details.

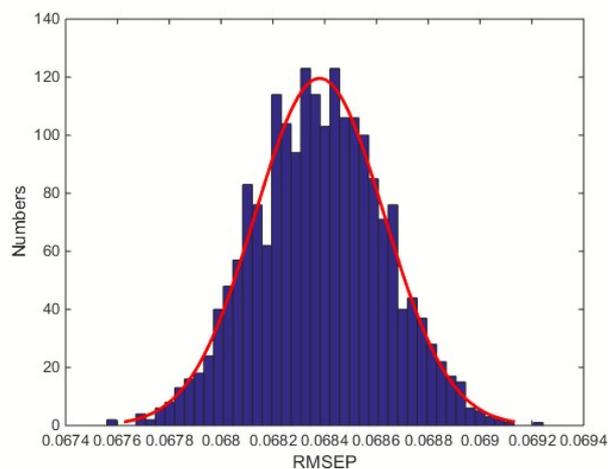


Fig.4 the distribution of RMSEP of tobacco dataset with 40 standardization samples with different hidden nodes numbers and different weights and biases.

Select activation function

The activation function in ELM is usually chosen as sigmoid and tanh function. They can be described mathematically with equations as follows:

Tanh function:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (11)$$

Sigmoid function:

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (12)$$

As tanh function has stronger gradients and can avoid bias in the gradients and is completely symmetric.

So tanh function has better performance than sigmoid function in general^{24, 25}. By comparing the RMSEP of different activation functions in TEAM on validation set for three spectral dataset, one can see from Table 2 that tanh activation function has significantly better performance than sigmoid function in most cases. Therefore, the tanh function has been chosen as activation function for calibration transfer in TEAM in this study.

Regularization for ELM to improve performance

In TEAM, the regularization term has been applied to improve generalization performance and make the solution more robust^{21, 26}. The regularization can be controlled by the parameter C in equation (7). As the results shown in Table.3, by adding a regularization term, the performance of TEAM is more robust with the growth of the number of subset samples. Also, the performance of TEAM is getting better. The regularization parameter C is optimized on validation sets. To optimize the regularization parameter C , a group of parameters were taken into consideration: 0, 100, 200, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 20000,

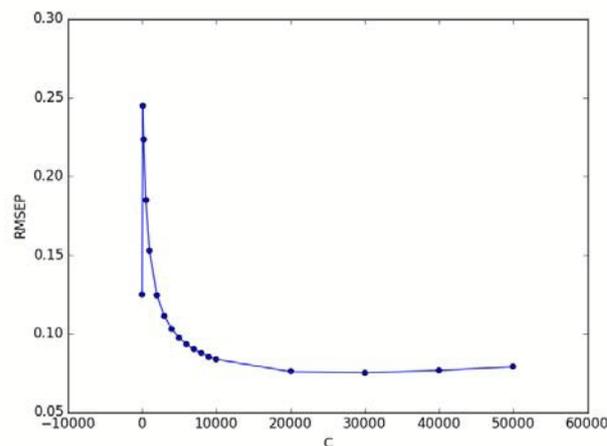


Fig.5 the variation trend of RMSEP with the change of regularization parameter C of corn (from M5 transfer to MP5) dataset with 20 standardization samples.

30000, 40000, 50000. Fig.5 shows the variation trend of RMSEP with the change of C for corn samples (from M5 transfer to MP5). With the growth of C , the RMSEP value is decrease and trend to be steady. From Fig S-15 to Fig S-16 in supporting information, the same result can be obtained by corn datasets (from MP6 transfer to MP5) and tobacco datasets. So, for these two datasets, we choose regularization parameter C as 50000. But for pharmaceutical tablets datasets, from the Fig S-17, we can find the RMSEP get minimize values when C is 500. So for pharmaceutical tablets datasets, we choose regularization parameter C as 500. But for pharmaceutical tablets datasets, the variation range of RMSEP on validation set is very small, so even choose regularization parameter C as 50000, we can still get good performance for TEAM method on independent test set. So for suggestion, when use TEAM, the default value of regularization parameter C can be set as 50000. But readers can also optimize the regularization parameter C based on the RMSEP value on validation set. Please see Fig S-15, Fig S-16 and Fig S-17 in the supplement information for more details.

Conclusions

In this study, the TEAM has been developed based on ELM-AE, and it was applied to transfer NIR datasets. The results obtained from three classic NIR datasets for calibration transfer show that TEAM is stable and can successfully correct the systematic differences between spectra obtained from different instruments. The performance of the proposed method, say TEAM, is better than PDS, GLS and CCA with three NIR datasets. TEAM can also achieve the best RMSEPs in most cases with small number of calibration set for establishing the transfer relationship. These advantages guarantee that TEAM will be an accurate and practical method to transfer spectra of slave instrument toward a well-established and maintained calibration model with few transfer samples, which eliminates the costly and time consuming recalibration.

Acknowledgements

This work is financially supported by the National Natural Science Foundation of China (Grants No. 21075138, Grants No. 21175157, Grants No. 21275164, Grants No. 21375151 and Grants No. 21305163), China Hunan Provincial science and technology department (Grants No. 2012FJ4139), Hunan Provincial Natural Science Foundation of China (Grants No. 14JJ3031), China Postdoctoral Science Foundation (No. 2014M552146) and National Instrumentation Program of China (No. 2011YQ03012407). The studies meet with the approval of the university's review board. We are grateful to all employees of this institute for their encouragement and support of this research.

Notes and references

1. S. Bureau, D. Ruiz, M. Reich, B. Gouble, D. Bertrand, J.-M. Audergon and C. M. Renard, *Food Chemistry*, 2009, **113**, 1323-1328.
2. N. De Belie, D. Pedersen, M. Martens, R. Bro, L. Munck and J. De Baerdemaeker, *Biosystems engineering*, 2003, **85**, 213-225.
3. J. A. Farrell, K. Higgins and J. H. Kalivas, *Journal of pharmaceutical and biomedical analysis*, 2012, **61**, 114-121.
4. A. R. Robinson and S. D. Mansfield, *The Plant Journal*, 2009, **58**, 706-714.
5. R. N. Feudale, N. A. Woody, H. Tan, A. J. Myles, S. D. Brown and J. Ferré, *Chemometrics and Intelligent Laboratory Systems*, 2002, **64**, 181-192.
6. T. Dean and T. Isaksson, *NIR news*, 1993, **4**, 14-15.
7. P. Geladi, D. MacDougall and H. Martens, *Applied spectroscopy*, 1985, **39**, 491-500.
8. T. B. Blank, S. T. Sum, S. D. Brown and S. L. Monfre, *Analytical chemistry*, 1996, **68**, 2987-2995.
9. S. Wold, H. Antti, F. Lindgren and J. Öhman, *Chemometrics and Intelligent Laboratory Systems*, 1998, **44**, 175-185.
10. H. Martens, M. Høy, B. M. Wise, R. Bro and P. B. Brockhoff, *Journal of Chemometrics*, 2003, **17**, 153-165.
11. E. Bouveresse, C. Hartmann, D. Massart, I. Last and K. Prebble, *Analytical Chemistry*, 1996, **68**, 982-990.
12. Y. Wang, D. J. Veltkamp and B. R. Kowalski, *Analytical chemistry*, 1991, **63**, 2750-2756.
13. J. S. Shenk and M. O. Westerhaus, *Journal*, 1989.
14. M. Forina, G. Drava, C. Armanino, R. Boggia, S. Lanteri, R. Leardi, P. Corti, P. Conti, R. Giangiacomo and C. Galliena, *Chemometrics and Intelligent Laboratory Systems*, 1995, **27**, 189-203.
15. W. Fan, Y. Liang, D. Yuan and J. Wang, *Analytica chimica acta*, 2008, **623**, 22-29.
16. G.-B. Huang, H. Zhou, X. Ding and R. Zhang, *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 2012, **42**, 513-529.
17. G.-B. Huang, X. Ding and H. Zhou, *Neurocomputing*, 2010, **74**, 155-163.
18. E. Cambria, G.-B. Huang, L. L. C. Kasun, H. Zhou, C. M. Vong, J. Lin, J. Yin, Z. Cai, Q. Liu and K. Li, *Intelligent Systems, IEEE*, 2013, **28**, 30-59.
19. G.-B. Huang, Q.-Y. Zhu and C.-K. Siew, *Neurocomputing*, 2006, **70**, 489-501.
20. G.-B. Huang, L. Chen and C.-K. Siew, *Neural Networks, IEEE Transactions on*, 2006, **17**, 879-892.
21. A. E. Hoerl and R. W. Kennard, *Technometrics*, 1970, **12**, 55-67.
22. W. B. Johnson and J. Lindenstrauss, *Contemporary mathematics*, 1984, **26**, 1.
23. R. W. Kennard and L. A. Stone, *Technometrics*, 1969, **11**, 137-148.
24. G. B. Orr and K.-R. Müller, *Neural networks: tricks of the trade*, Springer, 2003.
25. B. L. Kalman and S. C. Kwasny, 1992.
26. A. E. Hoerl and R. W. Kennard, *Technometrics*, 1970, **12**, 69-82.