# Investigating Intensity and Transversal Drift in Hyperspectral Imaging Data

Valerie Vaquet<sup>1</sup>, Patrick Menz<sup>2\*</sup>, Udo Seiffert<sup>2</sup>, and Barbara Hammer<sup>1</sup>

 Machine Learning Group Bielefeld University, Bielefeld - Germany
Cognitive Processes and Systems, Fraunhofer Institute of Factory Operation and Automation (IFF), Magdeburg - Germany

**Abstract**. When measuring data with hyperspectral cameras drift in the data distribution occurs over time and when the sensing device is changed. Frequently, this drift is a combination of intensity and wavelength shifts. In this contribution, we demonstrate that transfer component analysis together with subsampling constitutes a particular efficient and simple technology for spectral offset elimination which is applied to avoid the negative impact of drift on the classification performance. We demonstrate that this approach performs on par or better in comparison to established methods, and we also provide a theoretical motivation why this technology can deal with both, intensity as well as wavelength shift provided bounds on the smoothness of the functional data are given.

## 1 Introduction

Hyperspectral imaging is a sensing technology frequently used in quality control in food production and pharmaceutical applications, agriculture, environmental analysis and medical diagnosis [1]. A challenge in functional data analysis consists in the question how to achieve a suitable representation of the underlying functional signal as (often high-dimensional) vectors which are obtained by finite sampling. Different approaches for functional data analysis exist such as an adaptation of vectorial techniques to the underlying functional properties, specific metrics for high-dimensional data, expansion of the signals along suitable base functions, or a regularization of machine learning techniques for functional data analysis [2, 3, 4]. In this contribution, we will focus on signals from hyperspectral cameras for classification tasks using different approaches.

In practical applications, one core challenge of such sensor technology is the fact that, while devices are sophisticated, they are subject to changes in the environment and its physical components, which might alter over time. This causes drift in the distribution of the data collected at different times or across different instruments. In the literature, one way to efficiently deal with this challenge is offered by transfer learning or dedicated shift-invariant representations, which allow a transfer of an existing classification model to new settings based on few and possibly unlabeled data [5]. For hyperspectral cameras in the short-wavelength infrared (SWIR) and the visible near-infrared (VNIR) division of the electromagnetic spectrum, drift can be categorized into transversal and vertical shifts. While the measured wavelengths change when the components of a sensor age over time, introducing transversal shifts to the data distribution, vertical shifts are caused by changes in the measured intensities. They can be

<sup>\*</sup>Funding in the frame of the BMBF project TiM, 05M20PBA, and 05M20AFA is gratefully acknowledged.

caused by changes in lighting or by a different sensitivity of another sensor. In real world applications both shifts are observed simultaneously.

In this contribution, we investigate the shift characteristics individually to gain a better understanding on how to tackle them when building robust models. For this purpose, we create realistic data sets underlying only one type of drift in different severities based on real data and observed drift. We investigate and implement the objective which is also underlying transfer component analysis [6, 7] for the specific settings and domains and evaluate its performance in comparison to a representation based on a functional base as well as a classical standardization methods from chemometrics [8] for different application settings. The (somewhat surprising) superior performance of the first, comparably simple technology can be substantiated by mathematical bounds.

# 2 Approaches for shift-invariant spectral data processing

Spectral data representation. Spectral data have a functional form  $X : \mathbb{R} \to \mathbb{R}$ , mapping wavelengths to intensities, whereby typically smooth functions are considered, e.g. bounds on their modulus of continuity hold. In practice, spectra are typically observed at a finite number of frequencies only, resulting in possibly high-dimensional vectors  $x_i \in \mathbb{R}^d$  where components of  $x_i$  correspond to observations  $X(L) := (X(l_1), \ldots, X(l_d))$  at wavelength  $l_i$ . Typically, d is large, such that a compact representation or further regularization is necessary.

One common way to represent high-dimensional functional data is an approximation by a sum of spectral basis functions. One successful approach relies on Chebyshev polynomials  $T_j$  as base functions [4]. In this setting, each *d*-dimensional feature vector  $x_i \in \mathbb{R}^d$  is represented by d' coefficients  $c_{ij}$  of its approximation based on the first d' Chebyshev polynomials  $x_i \approx \sum_{j=0}^{d'} c_{ij}T_j(x_i)$  with  $T_j$  as the *j*th Chebyshev polynomial and ' $\approx$ ' referring to a least squares error, for example. Details can be found in [9]. A second possibility to reduce the dimension of (smooth) hyperspectral data is to apply subsampling, i.e.  $x_i \in \mathbb{R}^d$  is mapped to a lower dimensional space by projecting to a subset of the features at equally spaced intervals. These preprocessing steps have the effect that data can be represented by lower dimensional vectors, such that a better generalization ability and efficient processing can be expected for subsequent classification.

Transfer learning for different types of shift. Two types of shift can be present for spectral data X: (i) intensity/vertical shift, which corresponds to a function  $S_v : \mathbb{R} \to \mathbb{R}$  which is added to each spectrum X, yielding signals of the form  $X + S_v$  with finite-dimensional observation vector  $(X + S_v)(L) =$  $(X(l_1)+S_v(l_1),\ldots,X(l_d)+S_v(l_d)))$ , and (ii) wavelength/transversal shift, which corresponds to a strictly monotonic function  $S_h : \mathbb{R} \to \mathbb{R}$  which affects the wavelengths, i.e. the domain X, yielding the signal  $S_h \circ X$  with finite-dimensional observation vector  $X(S_h(L)) = (X(S_h(l_1)),\ldots,X(S_h(l_d)))$ . In real life data sets, shifts between the distributions of different sensors and over time are expected.

One classical way in the domain of chemometrics to deal with such shift is offered by standardization methods such as piecewise direct standardization (PDS) [8, 10]. Essentially, an explicit transformation matrix is determined by matching values of the original 'master' signal to signals of the transferred 'slave' corresponding to few neighbored wavelengths. As an alternative to PDS, it has been shown experimentally in [9] that a representation of spectra by Chebyshev polynomials together with subsequent z-transformation is robust to shift in the data, i.e. a transfer can be done immediately based on this representation. We will have a closer look at these two approaches in our experimental part.

In addition, we consider the behavior of a special instantiation of transfer component analysis (TCA) as introduced in the work [6], which, up to our knowledge, has not been used for transfer learning in the domain of spectral data analysis yet. Assume two data distributions corresponding to the original data  $x_i$  and transferred ones  $y_i$  are present. Assume a feature map  $\Phi$  is associated to the possibly nonlinear model (e.g. corresponding to a specific kernel). TCA aims for an optimization of the maximum mean discrepancy (MMD) as an important characteristic of the difference of two distributions with  $n_1$  data  $x_i$  and  $n_2$  data  $y_i$  respectively:  $|\frac{1}{n_1}\sum_i \Phi(x_i) - \frac{1}{n_2}\sum_i \Phi(y_i)|$ . Due to the high-dimensionality of spectral data, models can often be chosen as linear ones in this domain. In this case,  $\Phi$  is the identity and minimizing MMD can be done exactly by an elimination of the offsets: standardizing each data set component-wise is obtained by subtracting the mean spectrum  $\frac{1}{n_1}\sum_i x_i$  from  $x_i$  and  $\frac{1}{n_2}\sum_i y_i$  from  $y_i$ , respectively. We refer to this procedure as offset elimination in the following.

We can show that offset elimination is a particularly effective technology to deal with vertical drift. In addition, the smoothness of spectra accounts for a robustness to horizontal drift. More precisely, the following observation holds:

**Theorem 1** Assume a classifier f for spectra X is given, assume  $Y = X + S_v$  constitutes a vertical shift. If data are preprocessed with offset elimination, the accuracy of f on X and Y is the same.

We denote a horizontal shift  $S_h$  as limited by  $\epsilon$  iff  $|S_h(l) - l| \leq \epsilon$  for all wavelengths l. Assume the classifier f has sample-margin  $\eta$  on X. Assume f is uniformly continuous such that f differs by at most  $\eta$  if inputs differ by at most  $\eta_1$ . Assume X consists of uniformly continuous spectra where  $|X(l_1) - X(l_2)| \leq \eta_1/2$  for  $|l_1 - l_2| \leq \epsilon$ . Assume horizontal shift is limited by this value  $\epsilon$ . Then the result of f is the same for X and Y after offset elimination.

*Proof:* Assume the number of spectra is n. The first part follows from the observation that offset elimination of spectra with vertical shift yields  $(x_i + S_v)(l) - \sum_j (x_j + S_v)(l)/n = x_i(l) - \sum_j x_j(l)/n$  hence spectra stay invariant under vertical shift after offset elimination.

The second part follows from the estimation  $|(x_i(l) - \sum_j x_j(l)/n) - (y_i(l) - \sum_j y_j(l)/n)| \le |x_i(l) - x_i(S_h(l))| + \sum_j |x_j(l) - x_j(S_h(l))|/n \le \eta_1/2 + \eta_1/2 = \eta_1$  for limited horizontal shift. Hence the difference of the function values of these spectra after offset elimination is smaller than the margin  $\eta$ , i.e. the classification prescription is the same.

### **3** Experiments

*Datasets.* We use the following data measured by hyperspectral cameras in the SWIR and the VNIR division of the electromagnetic spectrum:

The SWIR data set contains the hyperspectral signature of 1500 pixels of Arabica, Robusta and immature Arabica coffee beans for each training and testing with balanced class distribution. For each pixel intensities at 249 wavelengths in the range of 1008.85nm and 2496.22nm are measured by a SWIR320me hyperspectral camera from Norsk Elektro Optikk.

ESANN 2021 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 6-8 October 2021, i6doc.com publ., ISBN 978287587082-7. Available from http://www.i6doc.com/en/.



Fig. 1: Example spectra (black) with moderate artificial intensity (orange) and transversal shifts (blue).

For the VNIR data set we recorded metal plates colored with RAL3000 fire red colors by different manufacturers with a NEO Hyspex VNIR-1600 camera by the same manufacturer. The data set consists of 3000 samples, which are balanced for the 4 classes, for each training and testing. Each sample contains data for 158 wavelengths in the range of 410.99nm and 982.68nm.

These initial data sets are enriched by horizontal and vertical shift as follows:

- **Intensity Shift:** We know from observation that intensity shift is wavelength specific but smooth. We simulate shift by randomly drawing shifts from a standard normal distribution at every 10th wavelength and interpolating in between. We multiply this random noise by different factors for different shift strengths. As previous measurements indicate that the shift is either positive or negative for all wavelengths, we add the minimum of the random shift as an offset. A visualization can be found in Fig. 1.
- **Transversal Shift:** For both hyperspectral cameras, we use the re-calibration report from the manufacturer, which were created every four years, to compute a baseline transversal shift. We interpolate the original data and collect the intensities at the shifted wavelengths. We repeat this also for multiples of the documented shift to simulate severe (albeit possibly unreasonable) transversal shifts, see Fig. 1.
- Methods. We compare the performance of the different pipelines:
- **Chebyshev:** Chebyshev coefficients are computed by approximating each spectrum by Chebyshev polynomials up to degree 50 based on the Chebyshev nodes, see [9] for details.
- **Chebyshev, z-transform:** Additionally, a z-transformation is applied to the obtained coefficients.
- **Offset-Elimination:** Offset-Elimination with prior subsampling is performed as defined above. During subsampling, each 5th feature is taken on the SWIR data set and each third feature on the VNIR data set, resulting in 50/53 features per sample.

ESANN 2021 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 6-8 October 2021, i6doc.com publ., ISBN 978287587082-7. Available from http://www.i6doc.com/en/.



(c) Transversal shift, SWIR data set. (d) Transversal shift, VNIR data set.

Fig. 2: Mean transfer accuracies of logistic regression over 5 experimental runs on datasets with intensity and transversal shifts with increasing drift severities.

- **Subsampled:** As a baseline subsampling of the raw spectra with the same sampling rates as above is considered.
- **PDS:** We transform the data of the transfer data set by the previously defined PDS method. Hyperparameters are optimized for the respective data set.

For each of these methods, we repeat our experiments 5 times and report the average transfer accuracy for different drift severities. After trying out different regularization strategies and hyper-parameters for logistic regression, we ended up with a logistic regression classifier without penalty.

*Results.* In our experiments with intensity shift, we observe constant high transfer accuracies for offset elimination and the z-transformed Chebyshev coefficients over all tested shift severities (see Fig. 2a, 2b). In contrast, for the Chebyshev representation without z-transformation our model declines with increasing shifts and performs slightly worse than the subsampled spectra. The experiments confirm Theorem 1 and show that a representation by Chebyshev coefficients alone is not invariant against intensity shift. The results on the PDS representation are poor on the SWIR and similar to the baseline on VNIR data.

On transversal shift, all methods decline with increasing shift (see Fig. 2c, 2d), whereby the reported results deal with extremal settings, i.e. four times larger than natural shift. While the results for the representation by offset elimination and z-transformed Chebyshev decline slowly with increasing transversal shifts,

the basic Chebyshev and the subsampled spectral representation decline very fast. For the SWIR dataset, they reach a level of an accuracy of 0.33 which corresponds to random guessing at a mean absolute shift of about 10nm. Similar to the experiment with intensity shift, the accuracies obtained by the PDS representation are low for all shifts. On the VNIR dataset (Fig. 2d) the PDS representation yields better results than offset elimination and z-scored Chebyshev coefficients for severe transversal shifts. Still, for more realistic smaller shifts, the use of the latter representations results in higher accuracies.

Overall, subsampling with offset elimination seems a simple but suitable choice to represent the non-stationary functional spectral data for reasonable ranges of shift. Chebyshev representations require a z-transformation.

#### 4 Conclusion

In this work we evaluated shift on spectral data by distinguishing between wavelength and intensity shifts. We showed that methods incorporating an elimination of the offset are invariant against intensity shifts. Additionally, we found that they are robust to moderate transversal shifts. As intensity shifts are frequently more severe than wavelength shifts in the data we analyzed so far, offset elimination is a rather simple method to build robust pipelines for real world applications. Further work should focus on finding suitable methods for data which comes from low quality devices (such as smartphones).

#### References

- [1] Celio Pasquini. Near infrared spectroscopy: A mature analytical technique with new perspectives–a review. Anal. Chim. Acta, 1026:8–36, 2018.
- [2] Jane-Ling Wang, Jeng-Min Chiou, and Hans-Georg Müller. Functional data analysis. Annu. Rev. Stat. Appl., 3(1):257–295, 2016.
- [3] Erzsébet Merényi, William H. Farrand, James V. Taranik, and Timothy B. Minor. Classification of hyperspectral imagery with neural networks: comparison to conventional tools. *EURASIP J. Adv. Signal Process.*, 2014:71, 2014.
- [4] Friedrich Melchert, Gabriele Bani, Udo Seiffert, and Michael Biehl. Adaptive basis functions for prototype-based classification of functional data. *Neural Comput. Appl.*, 32(24):18213–18223, 2020.
- [5] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, H. Zhu, Hui Xiong, and Q. He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109:43-76, 2021.
- [6] Sinno Jialin Pan, Ivor W. Tsang, James T. Kwok, and Qiang Yang. Domain adaptation via transfer component analysis. *IEEE Trans. Neural Networks*, 22(2):199–210, 2011.
- [7] Jiayuan Huang, Arthur Gretton, Karsten Borgwardt, Bernhard Schölkopf, and Alex Smola. Correcting sample selection bias by unlabeled data. In B. Schölkopf, J. Platt, and T. Hoffman, editors, Adv Neural Inf Process Syst, volume 19. MIT Press, 2007.
- [8] Yongdong Wang, David Veltkamp, and Bruce Kowalski. Multivariate instrument standardization. Annals of Chemsitry, 63:2750–2756, 1991.
- [9] Friedrich Melchert, Udo Seiffert, and Michael Biehl. Polynomial approximation of spectral data in lvq and relevance learning. *Machine Learning Reports*, 3:25–32, 2015.
- [10] E. Bouveresse and D.L. Massart. Improvement of the piecewise direct standardization procedure for the transfer of nir spectra for multivariate calibration. *Chemom. Intell. Lab. Syst.*, 32:201–213, 1996.