9-21-2018 10:00 AM

# Tensor-based Hyperspectral Image Processing Methodology and its Applications in Impervious Surface and Land Cover Mapping

Boyu Feng
*The University of Western Ontario*

Supervisor
Wang, Jinfei
*The University of Western Ontario*

Graduate Program in Geography

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

© Boyu Feng 2018

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Geographic Information Sciences Commons, and the Remote Sensing Commons

# Abstract

The emergence of hyperspectral imaging provides a new perspective for Earth observation, in addition to previously available orthophoto and multispectral imagery. This thesis focused on both the new data and new methodology in the field of hyperspectral imaging. First, the application of the future hyperspectral satellite EnMAP in impervious surface area (ISA) mapping was studied. During the search for the appropriate ISA mapping procedure for the new data, the spectral unmixing based on nonnegative matrix factorization (NMF) achieved the best success. The simulated EnMAP image shows great potential in urban ISA mapping with over 85% accuracy.

Unfortunately, the NMF based on the linear algebra only considers the spectral information and neglects the spatial information in the original image. The recent wide interest of applying the multilinear algebra in computer vision sheds light on this problem and raised the idea of nonnegative tensor factorization (NTF). This thesis found that the NTF has more advantages over the NMF when work with medium- rather than the high-spatial-resolution hyperspectral image. Furthermore, this thesis proposed to equip the NTF-based spectral unmixing methods with the variations adopted from the NMF. By adopting the variations from the NMF, the urban ISA mapping results from the NTF were improved by ~2%.

Lastly, the problem known as the *curse of dimensionality* is an obstacle in hyperspectral image applications. The majority of current dimension reduction (DR) methods are restricted to using only the spectral information, when the spatial information is neglected. To overcome this defect, two spectral-spatial methods: patch-based and tensor-patch-based, were thoroughly studied and compared in this thesis. To date, the popularity of the two solutions remains in computer vision studies and their applications in hyperspectral DR are limited. The patch-based and tensor-patch-based variations greatly improved the quality of dimension-reduced hyperspectral images, which then improved the land cover mapping results from them. In addition, this thesis proposed to use an improved method

to produce an important intermediate result in the patch-based and tensor-patch-based DR process, which further improved the land cover mapping results.

## Keywords

# Co-Authorship Statement

This thesis was prepared according to the integrated-article layout designed by the Faculty of Graduate Studies at Western University, London, Ontario, Canada. All the work stated in this thesis including methodology development, experimental testing, data analysis, modeling and writing draft manuscripts for publication was carried out by the author under the supervision of Dr. Jinfei Wang. Versions of Chapters 2, 3, and 4 have been accepted, submitted, or in preparation as co-authored peer reviewed journal papers. The coauthors can be found in the publication list below. Dr. Jinfei Wang provided the original conception of the application of the nonnegative matrix factorization method in remote sensing image processing. She proposed the urban feature extraction from high/medium resolution hyperspectral images for her NSERC Grant. Dr. Wang contributed in the development and formulation of methodology ideas and helped in establishing experimental procedures. She also provided valuable comments, editing and revision on the manuscripts, financial support, and software/hardware/data.

1. Feng, B. & Wang, J. (in preparation). Patch-based and tensor-patch-based dimension reduction methods for hyperspectral images.

2. Feng, B. & Wang, J. (under $1^{st}$ round revision). Constrained nonnegative tensor factorization for spectral mixture analysis of hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters*.

3. Feng, B. & Wang, J. (in press). Evaluation of unmixing methods for impervious surface area extraction from simulated EnMAP imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.*

# Acknowledgments

Four years ago, I first came to Canada as a foreigner. Today I have made a lot of friends on the land far away from home and fulfilled my research responsibility in my Ph.D study. On this journey, I obtained a lot of help from different people, some of whom I may never have the chance to return the favor. Thus, I would like to express my gratitude to them here, because without their help I could not even think about finishing my thesis.

In the first place, I thank my parents who although not physically near me emotionally supported me to overcome the obstacles along the way. My advisor Dr. Jinfei Wang is the best mentor I can wish for, who assists me in both my study and my life in Canada. Her acute and open-minded research-thinking always inspires me. Her warm-hearted generosity gives me enormous courage along both of my research and life paths. Within her power, Dr. Jinfei Wang provides me great research opportunities and financial help, which I cannot thank her enough. I thank Dr. Ying Zhang from Canada Centre for Remote Sensing. Her trust in my research ability helps me to develop a comprehensive research project. Dr. Peter Ashmore from Geography Department and Dr. Imtiaz Shah from Engineering Department give me a lot of help in the design of my research. For their valuable advices and thoughtful ideas, I am very thankful. Also, I appreciate the suggestions provided by Dr. Livio Tornabene, from whom I learned to keep a down-to-earth research mind and never forget the most basic concepts.

I am thankful for the unrequired and timely help from my senior coworker, Dr. Chuiqing Zeng. His encouragement gives me great motivation in our research project. I am also obliged to my lab mates Chunhua Liao, Yang Song, Xiaodong Huang, Peter Crawford, Xiaoxuan Sun, and Matthew Roffey. They show me the most kindness and make the countlessly stressful office hours bright. We developed the most sincere friendship.

In the end, I have to thank the Geography Department in Western University for providing such a great opportunity for me to study here. The Geography Department provides me with great research resources, including powerful computing unit and

# Table of Contents

# List of Tables

# List of Figures

# List of Appendices

# List of Abbreviations

SMA: spectral mixture analysis.

DR: dimension reduction.

ISA: impervious surface area.

V-I-S: vegetation-impervious surface-soil.

PPI: pixel purity index.

ICA: independent component analysis.

NMF: nonnegative matrix factorization.

sNMF: nonnegative matrix factorization with the sparseness constraint.

vNMF: nonnegative matrix factorization with minimum the volume constraint.

rNMF: nonnegative matrix factorization with the robust nonlinear constraint.

HFC: Harsanyi–Farrand–Chang method.

SAD: spectral angle distance.

NTF: nonnegative tensor factorization.

CP decomposition: CANDECOMP/PARAFAC decomposition.

BTD: block term decomposition.

MVNTF: matrix-vector nonnegative tensor factorization.

sMVNTF: matrix-vector nonnegative tensor factorization with the sparseness constraint.

vMVNTF: matrix-vector nonnegative tensor factorization ith the minimum volume constraint.

rMVNTF: matrix-vector nonnegative tensor factorization with the robust nonlinear constraint.

PCA: principle component analysis.

MDS: multidimensional scaling.

$k$NN: $k$ nearest neighbors.

LLE: locally linear embedding.

NPE: neighborhood preserving embedding.

LE: Laplacian eigenmaps.

LPP: locality preserving projections.

TLPP: tensor locality preserving projections.

TNPE: tensor neighborhood preserving embedding.

IPD: image patch distance.

RCM: region covariance matrix.

WRCM: weighted region covariance matrix.

# 1 Introduction

## 1.1 Research context

The urban impervious surface area (ISA) [1] and land cover are significant environment indicators (Arnold Jr and Gibbons 1996). Strong relations between the ISA/land cover and the urban hydrologic cycle (Miller et al. 2014, Arnold Jr and Gibbons 1996, Li et al. 2009, Loperfido et al. 2014, Pauleit, Ennos and Golding 2005, Leopold 1968), urban microclimate (Yuan and Bauer 2007, Voogt and Oke 2003, Zhou et al. 2014, Carlson and Arthur 2000, Xian and Crane 2006), and urban biodiversity (Seto, Güneralp and Hutyra 2012, McKinney 2008) have been found. The increase in urban ISA decreases the surface infiltration rate, increases water runoff, and triggers high peak streamflow (Paul and Meyer 2001). The land cover governs the behavior of the water runoff from urbanized and industrial areas discharges conveying large amount of nutrients, metals, pesticides, and other contaminants to streams (Karn and Harada 2001, Leopold 1968). Some impacts of the ISA on the urban energy budget are that, compared to previous areas, ISA absorbs more short-wave radiation, impedes release of long-wave energy to atmosphere, and increase the long-wave radiation to surrounding environment (Zhou et al. 2014), causing the urban heat island (Xu 2010). The land cover leads to intra-urban microclimate differences (Buyantuyev and Wu 2010). The increase of ISA also contributes to the loss of habitat, biomass, and carbon storage, jeopardizing biodiversity and ecosystem productivity (Seto et al. 2012). In addition, it has been found that the lack of urban green space due to ISA increase is disadvantageous to human physical activity, psychological well-being and the general public health of urban residents (Wolch, Byrne and Newell 2014). Furthermore, in the age of rapid urbanization, the urban ISA and land cover experience intense changes over the short time. Thus, in order to achieve urban sustainability, it is important to obtain up-to-date knowledge about the composition and distribution of urban ISA and land cover distribution.

---

[1] Impervious surfaces are mainly artificial structures that are covered by impenetrable materials such as asphalt and concrete.

Prior to the appearance of remote sensing, ISA/land cover studies were relatively disordered. For a long time, the ISA/land cover data were collected by agencies at various governmental levels under different standards, with little-to-no communication existing among them. This fact resulted in a confusing situation that forbid duplicated or shared work even in situations where the data were collected under similar premises (Anderson 1976). Additionally, at that time, the ISA/land cover data were obtained through ground surveys, which was labor intensive and time-consuming. Since the development of remote sensors, the remote sensing data has gradually become the mostly chosen source of information of information for ISA/land cover study.

The spectrum of a material is a plot of the percentage of reflectance, or emissivity, across a range of wavelengths. As all materials reflect, emit, transmit and absorb electromagnetic radiation based on the inherent physical structure and chemical composition of the material and the wavelength of the radiation (Vagni 2007), the spectrum of each material is unique. Remote sensors are designed to identify such spectral signatures and perform land cover mapping. The design of hyperspectral imaging of hundreds of spectral bands outperforms the traditional panchromatic (1 band), orthochromatic (usually has 3 bands), and multispectral (usually has 4~10 bands) sensors by providing more detailed spectral signatures (Figure 1-1).
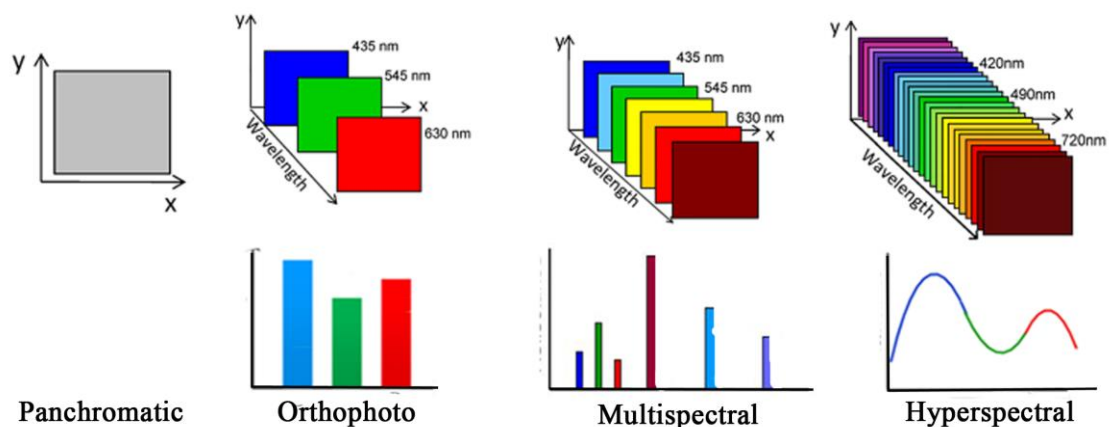


**Figure 1-1: Panchromatic/orthophoto/multispectral VS. hyperspectral images.**

In high demand, new hyperspectral sensors have been constantly designed. The near-future hyperspectral satellite project is Germany's environmental mapping and analysis program (EnMAP). Studies are needed for the future hypersepctral satellite, in order to take advantage of its new features. Depending on the new features and the practical objectives, different image processing procedures can be tested to find the appropriate solutions for the new hypersepctral satellite in different situations. This thesis focused on the ISA mapping ability of the future EnMAP image.

As the information in high dimensional data (e.g. hyperspectral images) are often redundant and contaminated by noise (Zhuo, Cheng and Zhang 2014), the efficient ISA/land cover mapping requires powerful processing methods. The majority of the current hyperspectral processing methods are applied in absence of the use of spatial information. Based on linear algebra, the spatial coordinates of pixels are often reconstructed into vectorized index during the process, which converts the original hyperspectral 3D cube into a 2D matrix (Figure 1-2). From now on, the methods based on linear algebra are referred to as the matrix-based methods. Fortunately, from the late 2000s, multilinear algebra is widely being studied in computer vision. Although the application of multilinear algebra in remote sensing is still limited, this trend sheds light on the possibility of improving series of hyperspectral image processing methods by preserving the spatial information. From now on, the methods based on multilinear algebra are referred to as the tensor[2]-based methods. In addition, inspired by the different variations of the widely used matrix-based hyperspectral image processing methods, I proposed to adopt them in tensor-based methods, which may also improve the final results. This thesis focused on improving two typical hyperspectral problems: spectral unmixing and dimension reduction, by tensor-based methods.

---

[2] In mathematics, an $N$th-order tensor is a $N$-dimensional array ($N \geq 3$).
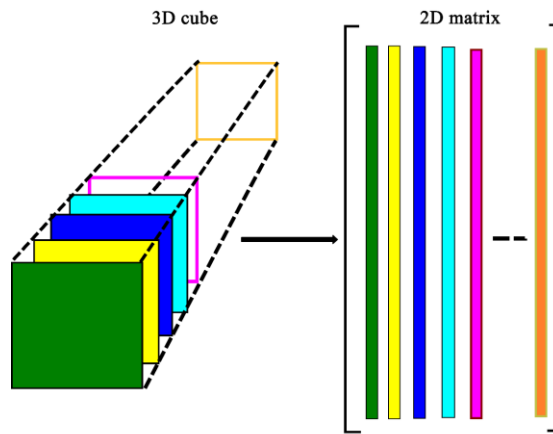
**Figure 1-2: Hyperspectral image matricization (3D cube to 2D matrix).**

## 1.2  Research objectives

The main objective of this thesis is to evaluate current, and develop new, remote sensing methodologies, on current/future hyperspectral images, in order to produce ISA/land cover maps with high accuracy and fidelity. Specifically, three sub-objectives for the three sub-studies are:

(1) To find the appropriate image processing methods for ISA mapping using the new simulated hyperspectral satellite EnMAP image.

(2) To upgrade one of the most robust spectral unmixing methods (nonnegative matrix factorization) from a matrix-based version to a tensor-based one and further improve the tensor-based method by incorporating additional constraints that have been previously added to the matrix-based method, to test improvement in the ISA mapping results.

(3) To upgrade the graph-based dimension reduction method to its tensor version and further improve the tensor-based method by the use of a new method for the intermediate results of adjacency graph/weight matrix, in the hope to improve the land cover mapping results.

Corresponding to the objectives, the following research questions are the focus of this thesis:

(1) Aiming to obtain accurate ISA maps, what processing procedures are appropriate for the new satellite EnMAP data when no additional data are available?

(2) How does the newly developed tensor-based spectral unmixing method assist the ISA mapping compared to the matrix-based method?

(3) Can the variations of the matrix-based spectral unmixing methods be implemented to the tensor-based methods? After this implementation, will the results of the tensor-based methods be improved?

(4) How does the newly developed spectral-spatial dimension reduction methods assist the land cover mapping compared to the matrix-based method?

(5) By improving the intermediate results (adjacency graph/weight matrix) of the tensor-based dimension reduction methods, can the final land cover mapping be improved?

## 1.3 Studied images

The quality of the final ISA and land cover maps relates to not only the discussed method, but also the studied hyperspectral images. Different processing methods may be appropriate solutions for different hyperspectral images. Features like spatial and spectral resolutions and locations (e.g. urban or suburban) of the hyperspectral images impact the outcomes for certain methods. Thus, one can seldom only work on image processing methods without considering the target. In order to comprehensively analyze the performances of studied methods regarding to different images, the hyperspectral images used for the analysis are carefully selected (Table 1-1). Firstly, this thesis provides spectral unmixing on one simulated EnMAP image (2013), whose results are valuable for the future application of the new hyperspectral sensor. The Copperas Cove, TX, Hydice image (1995) was also used in the spectral unmixing, for comparison purpose. As the EnMAP and Hydice images have different spatial and spectral resolution, they were able to reflect the different preferences of the spectral unmixing methods. In the dimension reduction experiments, the widely used Indian Pines AVIRIS image (1992) was used along with one newer CASI image (2013). The spectral band information of the EnMAP

and CASI data is provided in Appendix A. The spectral band information of the Hydice and AVIRIS data are open online (http://lesun.weebly.com/hyperspectral-data-set.html ).

**Table 1-1: Studied image.**

|  | Methods | Platforms | Location type | Spatial resolution | Spectral resolution | Temporal resolution |
|---|---|---|---|---|---|---|
| Simulated EnMAP | Spectral unmixing | Satellite | Urban-suburban | 30m | 88 bands (420-990nm) | 4 days |
| Hydice | Spectral unmixing | Airplane | Urban | 2m | 162 bands (400-2500nm) | NA |
| AVIRIS | Dimension reduction | Airplane | Agriculture | 20m | 200 bands (400-2500nm) | NA |
| CASI | Dimension reduction | Airplane | Urban | 1m | 72 bands (360-1050nm) | NA |

## 1.4  Background

### 1.4.1    ISA/land cover mapping

An ISA/land cover mapping literature search on Google Scholar, with approximately 80% - 90% of the articles being published in English, displayed literature from the 1990s to 2018. The majority of current ISA/land cover mapping used multispectral images, instead of hyperspectral images. This was largely due to a lack of hyperspectral sensors suitable for detecting and estimating various types of land cover, immature digital image processing techniques, and constrained computing power. In the 1990s, the number of publications on ISA/land cover mapping was less than 150, and they were mainly based on satellite imagery, e.g, Landsat TM (thematic mapper), Landsat MSS (multispectral scanner), SPOT, and AVHRR, with a small amount of publications referencing digitalized thematic maps and other ancillary data (e.g. population). At that time, the majority of ISA/land cover mapping literature was on the global scale. In the 2000s, the popularity of ISA/land cover mapping grew rapidly in the remote sensing community. More than 3000 publications focused on the application of remote sensing in ISA/land cover mapping. Greater amounts of remote sensing data became accessible for this task, including satellite missions (Landsat ETM+ (enhanced thematic mapper plus), ASTER,

MODIS, IKONOS, Quickbird, etc.); and airborne sensors (CASI, DAIS, etc.). Small-scale urban studies on country/city levels were available. In the 2010s, remote sensing continued to dominate the field of ISA/land cover mapping and the number of publications increased exponentially. Although the widely used data remained to be Landsat, MODIS, IKONOS, etc., various new classification methods were studied to improve the ISA/land cover mapping accuracy. In recent ISA/land cover studies, both per-pixel (Schneider, Friedl and Potere 2010) and spectral unmixing method (detailed explanation is in 1.4.3.1), were used with medium spatial resolution remote sensing images (e.g. Landsat and MODIS). The spectral mixture analysis was the most popular spectral unmixing method and several variations were proposed: normalized spectral mixture analysis, multiple endmember spectral mixture analysis, spatially adaptive spectral mixture analysis, etc. (Yang, Matsushita and Fukushima 2010, Deng and Wu 2013, Fan, Fan and Weng 2015). Other attempts with medium spatial resolution remote sensing often incorporated multi-temporal or multi-sensor remote sensing data (Lu, Moran and Hetrick 2011c, Lu et al. 2011b, Sung and Li 2012, Gao et al. 2012). With high spatial resolution images (e.g. IKONOS and Quickbird), object-oriented classification methods were popular (Hu and Weng 2011, Lu, Hetrick and Moran 2011a). In rough statistics, the overall classification accuracies ranged from 70% to 95% (Wickham et al. 2013, Lu et al. 2014, Zhang, Weng and Shao 2017, Chen et al. 2015, Momeni, Aplin and Boyd 2016), depending on the images used and methods used. Although hyperspectral imagery has not been widely used in ISA/land cover mapping, the continuous reflectance spectra of the hyperspectral remote sensing is superior to multispectral remote sensing by enabling detailed, precise mapping of earth surface compositions (Van der Meer et al. 2012). Further, experiments showed that hyperspectral images perform better in the low albedo areas (e.g. dark roofs and shadows) than multispectral images (Weng, Hu and Lu 2008).

## 1.4.2 Hyperspectral sensors

The emergence of the first hyperspectral sensor Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) in 1983 witnessed the beginning of the hyperspectral imaging era.

The hyperspectral sensors enable simultaneous data acquisition of hundreds or thousands of spectral bands, strengthening the conventional one-band panchromatic/infrared, three-band orthophoto and multispectral sensors and offering the information in the spectral ranges that are hidden from human eyes. Since then, a myriad of commercial airborne hyperspectral systems have been proposed: HyMap, Compact Airborne Spectral Imager (CASI), and digital airborne imaging spectrometer (DAIS) etc. In 2000, the launch of EO-1 Hyperion ended the absence of spaceborne hyperspectral sensors, arousing the development of a series spaceborne hyperspectral sensors (Buckingham and Staenz 2008). The compact high-resolution imaging spectrometer (CHRIS) designed by the European Space Agency (ESA) is another currently operating spaceborne hyperspectral sensor. It has been stated that the hyperspectral technology probably will be the future of remote sensing (Bioucas-Dias et al. 2013). The hyperspectral imagery has proved itself in urban ISA and land cover mapping (Fauvel et al. 2008, Benediktsson, Palmason and Sveinsson 2005, Huang and Zhang 2009).

## 1.4.3 Problems with hyperspectral image

### 1.4.3.1 Mixed pixel problem

For the hyperspectral images of coarse/medium spatial resolutions, the detailed ISA/land cover distribution can hardly be achieved by per-pixel analyses, since multiple land cover types co-exist in one pixel. This problem concerning the coarse/medium spatial resolution is often referred to as mixed pixel problem. The objective of the mixed pixel problem is to find an abundance map indicating the existing materials and their percentages in each pixel. To address this problem, spectral unmixing has been proposed in the early 1980s (Dozier 1981), and has since been widely adopted in coarse/medium hyperspectral imagery analyses (Powell et al. 2007, Roberts et al. 1998, Wu and Murray 2003, Liu et al. 2004). The spectral unmixing aims at unmixing the pixels by modeling the reflectance

value of the target pixel from more than one endmember[3]. A wide range of classification algorithms has been proposed on a subpixel level. The spectral mixture analysis (SMA) is one of the most commonly used spectral unmixing methods (Liu et al. 2004), thanks to its simplicity. The SMA assumes the reflectance of a given pixel equal to the sum of the reflectance of each material multiplied by its fraction within a pixel (Figure 1-3).
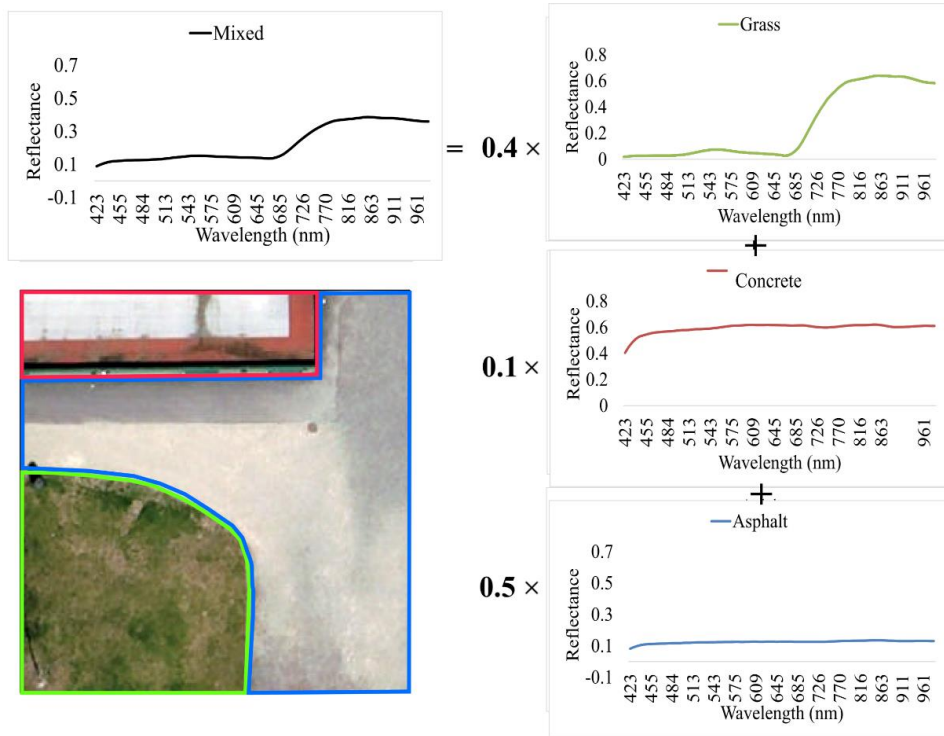


**Figure 1-3: Spectral mixture analysis illustration of one pixel.**

However, the SMA method requires the input of endmembers. Thus, a pre-processing step of endmember extraction is often involved. The process of endmember extraction is to build a reference spectral library. For a successful reference spectral library, it is necessary to have good representation of both groups within the library collection and its class on the ground (Powell et al. 2007). Two data sources are normally used for endmember extraction: field/lab spectrometer and remote sensing images. Using the

---

[3] An endmember is a pure spectrum that is chosen to represent pure surface materials in a spectral image.

spectrometer to measure the spectra of target materials in the field or lab is straightforward and a myriad of spectral libraries are built up using this method. The widely accepted USGS spectral library collects the endmember spectra in a lab, where researchers measure the pure material surface using four different spectrometers to cover the spectral range of 0.2 to 150 μm (Clark et al. 2007). Though lab-based endmember collection procedures have been established as the current standard, remote sensing images in natural settings may have different spectral signatures compared to the data collected on the ground or in the lab. Firstly, lab endmember data is seldom acquired under the same condition as the airborne or spaceborne data (Plaza et al. 2004). Secondly, the atmospheric effects that play a great role in airborne or spaceborne data cannot be revealed in lab data. Thus, a more precise method is to extract endmembers directly from the remote sensing imagery. Most of the time, the endmember by nature is small in number in a remote sensing image. As a result, their appearances are often anomalies, making it difficult to locate them (Chang et al. 2006).

More recently, a new solution for SMA has been proposed: nonnegative matrix factorization (NMF). The NMF simultaneously calculates the endmember and abundance maps, by applying linear algebra to decompose the original hyperspectral image into an endmember matrix and an abundance matrix. Stemming from computer vision studies, the original NMF passes through various modifications to accommodate the physical concepts in the SMA process.

### 1.4.3.2    Curse of dimensionality

The problem of the *curse of dimensionality*, which refers to the noise and redundancy in the high dimensional data, is an obstacle in the application of hyperspectral image. An appropriate dimension reduction (DR) process prepares the data for more effective information retrieval by revealing low-dimensional structures hidden in high-dimensional spaces. The DR methods can be roughly divided into linear and nonlinear. Generally, linear DR methods assume that the data lie close to a lower dimensional linear subspace and result in linear combinations of the original variables. Due to this simple implementation, the linear DR algorithms are well developed and embrace great

popularity. The principal component analysis (PCA) is one of the widely-used linear DR solutions. It is a non-targeting linear DR algorithm and can be applied to various datasets without complex modification. PCA was first proposed by Pearson in 1901 and has experienced several modifications to become the transformation we know today (Hotelling 1933, Wold 1968). PCA provides solutions to various multivariate problems, including data reduction and finding classes of similar objects (Wold, Esbensen and Geladi 1987). The multispectral or vector character of most remote sensing imagery enables it to transform the original spectral space to feature spaces constructed by new components (Richards and Richards 1999). The PCA is a statistical procedure that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components (PCs) (Figure 1-4).
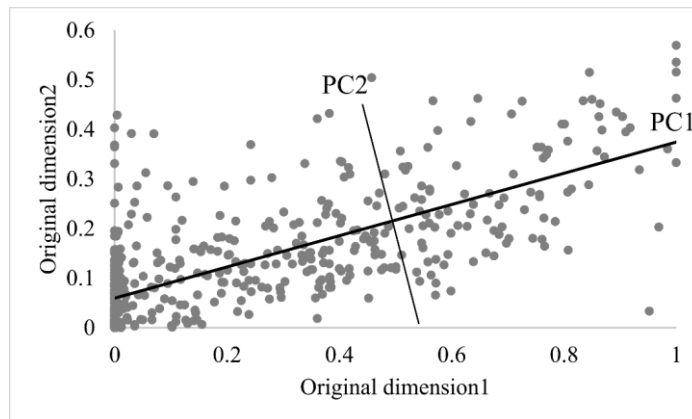


**Figure 1-4: Principal component analysis illustration.**

Yet, the results from linear DR can fail to be optimal due to the nonlinear features lying in a lot of data. For example, to linearly project data that is on or near a curved low dimensional space onto a linear subspace will lead to a large error (e.g. Swiss roll dataset (Figure 1-5)), unless the subspace has a higher dimension than the original curved space, which contradicts the objective of DR (Kambhatla and Leen 1993). In light of this, more complex nonlinear DR algorithms that address the underlying nonlinear features in these data have been proposed (e.g. locally linear embedding (LLE) and Laplacian eigenmaps (LE)). The nonlinear DR implies that the linear DR only focuses on preserving the

straight-line Euclidean distance of the data and neglects the nonlinear structure in the data (Zhang and Zha 2004). Two streams of nonlinear DR algorithms are based on geodesic distances and local structures, respectively. The first stream of nonlinear DR algorithms uses the geodesic distance to reveal the true data structure (Tenenbaum, De Silva and Langford 2000). The second stream of nonlinear DR algorithms maintains the local nonlinear structure during any transformation (Hastie and Stuetzle 1989).
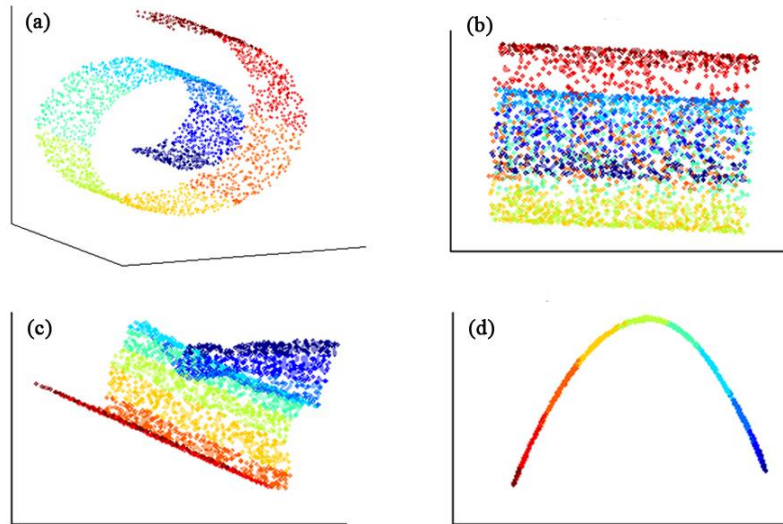


**Figure 1-5: Swiss roll dataset illustration:**

**(a) original Swiss roll dataset; dimension-reduced result from (b) PCA (principle component analysis); (c) LLE (locally linear embedding); and (d) LE (Laplacian eigenmaps).**

## 1.4.4    Multilinear algebra notation and preliminaries

The above descriptions of the two hyperspectral image processing methods (SMA (spectral mixture analysis) and DR (dimension reduction)) both have the potential to benefit from the current trend of multilinear algebra. Before introducing how to apply multilinear algebra in SMA and DR, I will first provide the basic concepts of multilinear algebra. In mathematics, an $N$th-order tensor is a $N$-dimensional array: $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$. In simple cases, a vector is a first-order tensor, a matrix is a second-order tensor, and a data cube (e.g. a hyperspectral image) is a third-order tensor. The order of a tensor is the

number of dimensions, which can be also called ways or modes. In this thesis, I maintained the widely used notations among mathematical articles. Vectors (first-order tensor) are denoted by boldface lowercase letters, e.g. $\mathbf{a}$, and their $i$th entry is denoted by $a_i$. Matrices (second-order tensor) are denoted by boldface capital letters, e.g., $\mathbf{A}$, and their element $(i, j)$ is denoted by $a_{ij}$. Third-order or higher-order tensors are denoted by boldface Euler script letters, e.g. $\boldsymbol{\mathcal{X}}$, and elements $(i, j, k)$ of a third-order tensor are denoted by $a_{ijk}$. Subscript indices range from 1 to their capital versions, e.g., $i = 1, \cdots, I$. The $n$th element in a sequence is denoted by a superscript in parentheses, e.g., $\boldsymbol{\mathcal{X}}^{(n)}$ denotes the $n$th tensor in a sequence. Subarrays are derived from fixing a subset of index. A colon is used to denote all elements of a mode. In the case of matrices, the subarrays are rows ($a_{i:}$) and columns ($a_{:j}$). For higher-order tensors, the subarrays are called fibers when fixing all but one indices or called slices when fixing all but two indices (Figure 1-6).
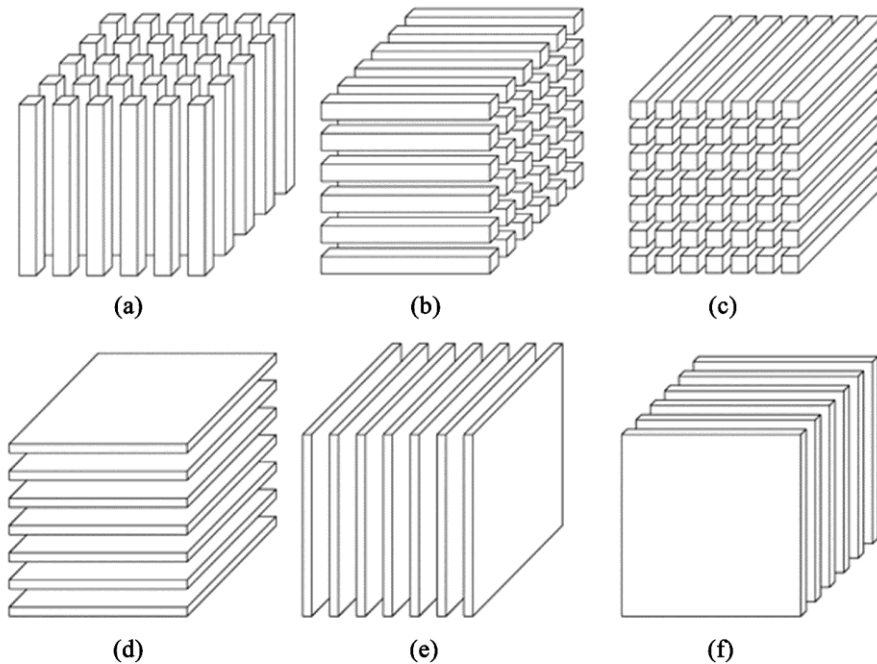


**Figure 1-6: Fibers and slices of a third-order tensor:**
**(a) Mode-1 fibers $a_{:jk}$; (b) Mode-2 fibers $a_{i:k}$; (c) Mode-3 fibers $a_{ij:}$; (d) Horizontal slices $X_{i::}$; (e) Lateral slices $X_{:j:}$; and (f) Frontal slices $X_{::k}$ (Kolda and Bader 2009).**

Unfolding is another basic transformation of a tensor, and is also referred to as matricization and flattening. The unfolding process reorders the elements of a tensor into a matrix, which is a step required by certain analyses. There are many ways to assemble the elements from a tensor into a matrix. This thesis only applied the mode-$n$ unfolding, denoted by $X_{(n)}$. The mode-$n$ unfolding takes the mode-$n$ fibers as columns in the resulting matrix.

Analogous to the matrix Frobenius norm denoted as $\| A \|$, the norm of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is the square root of the sum of the squares of all its elements:

$$\| \mathcal{X} \| = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_N=1}^{I_N} x_{i_1 i_2 \cdots i_N}^2}, \tag{1-1}$$

Tensors can be multiplied by matrices using the $n$-mode product. It is a basic tensor calculation that is used in tensor-based multilinear operations. The $n$-mode product multiplies a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_n \times \cdots \times I_N}$ by a matrix $U \in \mathbb{R}^{J \times I_n}$ in mode $n$, and is denoted by $\mathcal{X} \times_n U$. The product result has a size of $I_1 \times \cdots I_1 \times \cdots \times I_{n-1} \times J \times I_{n+1} \times \cdots \times I_N$. The element of the result is calculated as below:

$$(\mathcal{X} \times_n U)_{i_1 \cdots i_{n-1} j i_{n+1} \cdots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 i_2 \cdots i_N} u_{j i_n}, \tag{1-2}$$

The $n$-mode product can be also interpreted as multiplying each mode-n fiber with the matrix:

$$\mathcal{Y} = \mathcal{X} \times_n U \iff Y_{(n)} = U X_n \tag{1-3}$$

In order to explore the hidden information in a tensor, different tensor decomposition methods have been proposed to deal with different problems. Two main tensor decomposition methods include CANDECOMP/PARAFAC (CP) decomposition and Tucker decomposition. Hitchcock proposed the initial idea of CP decomposition in 1927 (Hitchcock 1927). CP decomposition did not become popular until 1970 when Carroll and Chang re-introduced the concept of CANDECOMP (canonical decomposition) (Carroll and Chang 1970) and Harshman re-introduced the concept of PARAFAC (parallel factors) (Harshman 1970) in the psychometrics community. The CP decomposition adopts a polyadic form, expressing a tensor as the sum of a finite number

of rank-one tensors (Figure 1-7). A rank-one tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$ is an $N$th-order tensor that can be obtained from the outer product of $N$ vectors. The outer product is the tensor product of two coordinate vectors. Given two tensors $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_p}$ and $\mathcal{B} \in \mathbb{R}^{J_1 \times J_2 \cdots \times J_Q}$, the outer product is written as $\mathcal{A} \circ \mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_p \times J_1 \times J_2 \cdots \times J_Q}$. The element is obtained as below:

$$(\mathcal{A} \circ \mathcal{B})_{i_1 i_2 \ldots i_P j_1 j_2 \ldots j_Q} = a_{i_1 i_2 \ldots i_P} b_{j_1 j_2 \ldots j_Q}, \tag{1-4}$$

The CP decomposition for a third-order tensor $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ can be written as:

$$\mathcal{X} \approx \sum_{r=1}^{R} a_r \circ b_r \circ c_r, \tag{1-5}$$

where $R$ is the number of dimensions of $\mathcal{X}$. The smallest R that fulfills the Equation (1-5) is the rank of a tensor. Unfortunately, there is no straightforward way to determine the rank of a tensor.

The Tucker decomposition was first proposed by Tucker in 1963 (Tucker 1963) (Figure 1-7). It decomposes a tensor into a core tensor multiplied by a matrix along each mode:

$$\mathcal{X} \approx \mathcal{G} \times_1 A \times_2 B \times_3 C, \tag{1-6}$$

The size of the core $\mathcal{G}$ is the key input in Tucker decompositions, which is not readily available but is obtained through experiments.
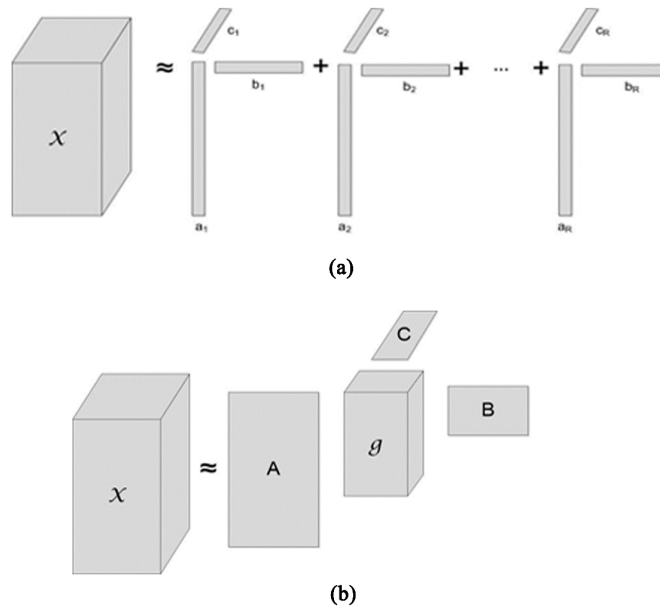
**Figure 1-7: Tensor decomposition illustration:**
**(a) CP (CANDECOMP/PARAFAC) decomposition; (b) Tucker decomposition.**

As the majority of current image processing methods are based on 2D matrix calculation, the hyperspectral 3D cube needs to be matricized in order to be processed, which discards the spatial information along the two spatial dimensions. The above-introduced multilinear algebra enables the direct calculation between 3D cubes that are referred to as 3D tensors. Thus, the tensor-based image processing method preserves the spatial information in the original image.

## 1.5 Organization of the thesis

The thesis consists of five chapters. Chapter 1 provides the research context, explains the thesis objectives, states the major research questions, and provides the necessary background. Chapters 2-4 separately covers:

(1) Evaluation of SMA methods for simulated EnMAP hyperspectral imagery.

(2) Constrained nonnegative tensor factorization for SMA of hyperspectral image.

(3) The application of spectral-spatial representation of hyperspectral images in DR.

Chapter 2 concerns the application of the future EnMAP image in ISA mapping. Chapter 3 concerns the application of the multilinear algebra in spectral unmixing for ISA mapping using hyperspectral images. Chapter 4 concerns the application of the multilinear algebra in dimension reduction for land cover mapping using hyperspectral images. Figure 1-8 shows the relations between Chapters 2-4. The ultimate achievement of them is to obtain accurate and reliable ISA/land cover maps from hyperspectral images. Chapter 5 concludes the main findings from Chapters 2-4.
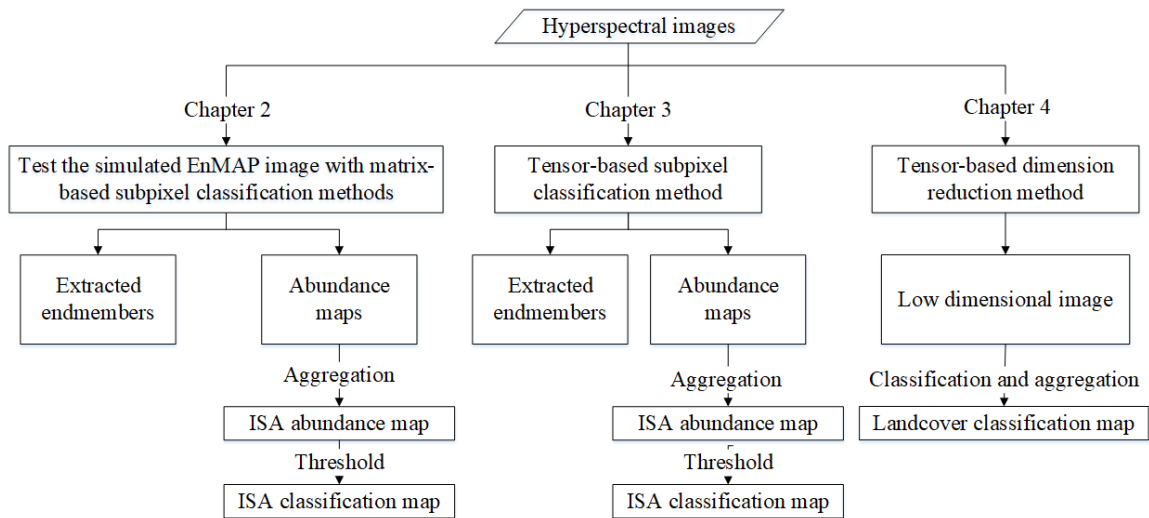


**Figure 1-8:  Relationship among the three sub-studies.**

## References

Anderson, J. R. 1976. *A land use and land cover classification system for use with remote sensor data*. US Government Printing Office.

Arnold Jr, C. L. & C. J. Gibbons (1996) Impervious surface coverage: the emergence of a key environmental indicator. *Journal of the American planning Association,* 62**,** 243-258.

Benediktsson, J. A., J. A. Palmason & J. R. Sveinsson (2005) Classification of hyperspectral data from urban areas based on extended morphological profiles. *Geoscience and Remote Sensing, IEEE Transactions on,* 43**,** 480-491.

Bioucas-Dias, J. M., A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi & J. Chanussot (2013) Hyperspectral remote sensing data analysis and future challenges. *Geoscience and Remote Sensing Magazine, IEEE,* 1**,** 6-36.

Buckingham, R. & K. Staenz (2008) Review of current and planned civilian space hyperspectral sensors for EO. *Canadian Journal of Remote Sensing,* 34**,** S187-S197.

Buyantuyev, A. & J. Wu (2010) Urban heat islands and landscape heterogeneity: linking spatiotemporal variations in surface temperatures to land-cover and socioeconomic patterns. *Landscape ecology,* 25**,** 17-33.

Carlson, T. N. & S. T. Arthur (2000) The impact of land use—land cover changes due to urbanization on surface microclimate and hydrology: a satellite perspective. *Global and planetary change,* 25**,** 49-65.

Carroll, J. D. & J.-J. Chang (1970) Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika,* 35**,** 283-319.

Chang, C.-I., C.-C. Wu, W.-m. Liu & Y.-C. Ouyang (2006) A new growing method for simplex-based endmember extraction algorithm. *Geoscience and Remote Sensing, IEEE Transactions on,* 44**,** 2804-2819.

Chen, J., J. Chen, A. Liao, X. Cao, L. Chen, X. Chen, C. He, G. Han, S. Peng & M. Lu (2015) Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing,* 103**,** 7-27.

Clark, R. N., G. A. Swayze, R. Wise, K. E. Livo, T. M. Hoefen, R. F. Kokaly & S. J. Sutley. 2007. USGS digital spectral library splib06a. US Geological Survey Reston, VA.

Deng, C. & C. Wu (2013) A spatially adaptive spectral mixture analysis for mapping subpixel urban impervious surface distribution. *Remote Sensing of Environment,* 133**,** 62-70.

Dozier, J. (1981) A method for satellite identification of surface temperature fields of subpixel resolution. *Remote Sensing of environment,* 11**,** 221-229.

Fan, F., W. Fan & Q. Weng (2015) Improving urban impervious surface mapping by linear spectral mixture analysis and using spectral indices. *Canadian Journal of Remote Sensing,* 41**,** 577-586.

Fauvel, M., J. A. Benediktsson, J. Chanussot & J. R. Sveinsson (2008) Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. *Geoscience and Remote Sensing, IEEE Transactions on,* 46**,** 3804-3814.

Gao, F., E. B. de Colstoun, R. Ma, Q. Weng, J. G. Masek, J. Chen, Y. Pan & C. Song (2012) Mapping impervious surface expansion using medium-resolution satellite image time series: a case study in the Yangtze River Delta, China. *International Journal of Remote Sensing,* 33**,** 7609-7628.

Harshman, R. A. (1970) Foundations of the PARAFAC procedure: Models and conditions for an" explanatory" multimodal factor analysis.

Hastie, T. & W. Stuetzle (1989) Principal curves. *Journal of the American Statistical Association,* 84**,** 502-516.

Hitchcock, F. L. (1927) The expression of a tensor or a polyadic as a sum of products. *Studies in Applied Mathematics,* 6**,** 164-189.

Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components. *Journal of educational psychology,* 24**,** 417.

Hu, X. & Q. Weng (2011) Impervious surface area extraction from IKONOS imagery using an object-based fuzzy method. *Geocarto International,* 26**,** 3-20.

Huang, X. & L. Zhang (2009) A comparative study of spatial approaches for urban mapping using hyperspectral ROSIS images over Pavia City, northern Italy. *International Journal of Remote Sensing,* 30**,** 3205-3221.

Kambhatla, N. & T. K. Leen. 1993. Fast nonlinear dimension reduction. In *Neural Networks, 1993., IEEE International Conference on*, 1213-1218. IEEE.

Karn, S. K. & H. Harada (2001) Surface water pollution in three urban territories of Nepal, India, and Bangladesh. *Environmental Management,* 28**,** 483-496.

Kolda, T. G. & B. W. Bader (2009) Tensor decompositions and applications. *SIAM review,* 51**,** 455-500.

Leopold, L. B. (1968) Hydrology for urban land planning: A guidebook on the hydrologic effects of urban land use.

Li, H., L. J. Sharkey, W. F. Hunt & A. P. Davis (2009) Mitigation of impervious surface hydrology using bioretention in North Carolina and Maryland. *Journal of Hydrologic Engineering,* 14**,** 407-415.

Liu, W., K. C. Seto, E. Y. Wu, S. Gopal & C. E. Woodcock (2004) ART-MMAP: A neural network approach to subpixel classification. *Geoscience and Remote Sensing, IEEE Transactions on,* 42**,** 1976-1983.

Loperfido, J. V., G. B. Noe, S. T. Jarnagin & D. M. Hogan (2014) Effects of distributed and centralized stormwater best management practices and land cover on urban stream hydrology at the catchment scale. *Journal of Hydrology,* 519**,** 2584-2595.

Lu, D., S. Hetrick & E. Moran (2011a) Impervious surface mapping with Quickbird imagery. *International journal of remote sensing,* 32**,** 2519-2533.

Lu, D., G. Li, W. Kuang & E. Moran (2014) Methods to extract impervious surface areas from satellite images. *International Journal of Digital Earth,* 7**,** 93-112.

Lu, D., G. Li, E. Moran, M. Batistella & C. C. Freitas (2011b) Mapping impervious surfaces with the integrated use of Landsat Thematic Mapper and radar data: A case study in an urban–rural landscape in the Brazilian Amazon. *ISPRS Journal of Photogrammetry and Remote Sensing,* 66**,** 798-808.

Lu, D., E. Moran & S. Hetrick (2011c) Detection of impervious surface change with multitemporal Landsat images in an urban–rural frontier. *ISPRS Journal of Photogrammetry and Remote Sensing,* 66**,** 298-306.

McKinney, M. L. (2008) Effects of urbanization on species richness: a review of plants and animals. *Urban ecosystems,* 11**,** 161-176.

Miller, J. D., H. Kim, T. R. Kjeldsen, J. Packman, S. Grebby & R. Dearden (2014) Assessing the impact of urbanization on storm runoff in a peri-urban catchment using historical change in impervious cover. *Journal of Hydrology,* 515**,** 59-70.

Momeni, R., P. Aplin & D. S. Boyd (2016) Mapping complex urban land cover from spaceborne imagery: the influence of spatial resolution, spectral band set and classification approach. *Remote Sensing,* 8**,** 88.

Paul, M. J. & J. L. Meyer (2001) Streams in the urban landscape. *Annual Review of Ecology and Systematics***,** 333-365.

Pauleit, S., R. Ennos & Y. Golding (2005) Modeling the environmental impacts of urban land use and land cover change—a study in Merseyside, UK. *Landscape and urban planning,* 71**,** 295-310.

Plaza, A., P. Mart ínez, R. Pérez & J. Plaza (2004) A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data. *Geoscience and Remote Sensing, IEEE Transactions on,* 42**,** 650-663.

Powell, R. L., D. A. Roberts, P. E. Dennison & L. L. Hess (2007) Sub-pixel mapping of urban land cover using multiple endmember spectral mixture analysis: Manaus, Brazil. *Remote Sensing of Environment,* 106**,** 253-267.

Richards, J. A. & J. Richards. 1999. *Remote sensing digital image analysis*. Springer.

Roberts, D. A., M. Gardner, R. Church, S. Ustin, G. Scheer & R. Green (1998) Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. *Remote Sensing of Environment,* 65**,** 267-279.

Schneider, A., M. A. Friedl & D. Potere (2010) Mapping global urban areas using MODIS 500-m data: New methods and datasets based on 'urban ecoregions'. *Remote Sensing of Environment,* 114**,** 1733-1746.

Seto, K. C., B. Güneralp & L. R. Hutyra (2012) Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences,* 109**,** 16083-16088.

Sung, C. Y. & M.-H. Li (2012) Considering plant phenology for improving the accuracy of urban impervious surface mapping in a subtropical climate regions. *International journal of remote sensing,* 33**,** 261-275.

Tenenbaum, J. B., V. De Silva & J. C. Langford (2000) A global geometric framework for nonlinear dimensionality reduction. *Science,* 290**,** 2319-2323.

Tucker, L. R. (1963) Implications of factor analysis of three-way matrices for measurement of change. *Problems in measuring change,* 15**,** 122-137.

Vagni, F. (2007) Survey of hyperspectral and multispectral imaging technologies.

Van der Meer, F. D., H. M. A. Van der Werff, F. J. Van Ruitenbeek, C. A. Hecker, W. H. Bakker, M. F. Noomen, M. Van Der Meijde, E. J. M. Carranza, J. B. De Smeth & T. Woldai (2012) Multi-and hyperspectral geologic remote sensing: A review. *International Journal of Applied Earth Observation and Geoinformation,* 14**,** 112-128.

Voogt, J. A. & T. R. Oke (2003) Thermal remote sensing of urban climates. *Remote sensing of environment,* 86**,** 370-384.

Weng, Q., X. Hu & D. Lu (2008) Extracting impervious surfaces from medium spatial resolution multispectral and hyperspectral imagery: a comparison. *International Journal of Remote Sensing,* 29**,** 3209-3232.

Wickham, J. D., S. V. Stehman, L. Gass, J. Dewitz, J. A. Fry & T. G. Wade (2013) Accuracy assessment of NLCD 2006 land cover and impervious surface. *Remote Sensing of Environment,* 130**,** 294-304.

Wolch, J. R., J. Byrne & J. P. Newell (2014) Urban green space, public health, and environmental justice: The challenge of making cities 'just green enough'. *Landscape and Urban Planning,* 125**,** 234-244.

Wold, H. O. A. 1968. *Nonlinear estimation by iterative least square procedures.*

Wold, S., K. Esbensen & P. Geladi (1987) Principal component analysis. *Chemometrics and intelligent laboratory systems,* 2**,** 37-52.

Wu, C. & A. T. Murray (2003) Estimating impervious surface distribution by spectral mixture analysis. *Remote sensing of Environment,* 84**,** 493-505.

Xian, G. & M. Crane (2006) An analysis of urban thermal characteristics and associated land cover in Tampa Bay and Las Vegas using Landsat satellite data. *Remote Sensing of environment,* 104**,** 147-156.

Xu, H. (2010) Analysis of impervious surface and its impact on urban heat environment using the normalized difference impervious surface index (NDISI). *Photogrammetric Engineering & Remote Sensing,* 76**,** 557-565.

Yang, F., B. Matsushita & T. Fukushima (2010) A pre-screened and normalized multiple endmember spectral mixture analysis for mapping impervious surface area in Lake Kasumigaura Basin, Japan. *ISPRS Journal of Photogrammetry and Remote Sensing,* 65**,** 479-490.

Yuan, F. & M. E. Bauer (2007) Comparison of impervious surface area and normalized difference vegetation index as indicators of surface urban heat island effects in Landsat imagery. *Remote sensing of Environment,* 106**,** 375-386.

Zhang, L., Q. Weng & Z. Shao (2017) An evaluation of monthly impervious surface dynamics by fusing Landsat and MODIS time series in the Pearl River Delta, China, from 2000 to 2015. *Remote Sensing of Environment,* 201**,** 99-114.

Zhang, Z.-y. & H.-y. Zha (2004) Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *Journal of Shanghai University (English Edition),* 8**,** 406-424.

Zhou, W., Y. Qian, X. Li, W. Li & L. Han (2014) Relationships between land cover and the surface urban heat island: seasonal variability and effects of spatial and thematic resolution of land cover data on predicting land surface temperatures. *Landscape ecology,* 29**,** 153-167.

Zhuo, L., B. Cheng & J. Zhang (2014) A comparative study of dimensionality reduction methods for large-scale image retrieval. *Neurocomputing,* 141**,** 202-210.

# 2   Evaluation of unmixing methods for simulated EnMAP hyperspectral imagery

The distribution of impervious surface area (ISA) is an important input in a wide range of urban ecosystem studies, including urban hydrology, urban climate, land use planning, and resource management (Arnold Jr and Gibbons 1996, Voogt and Oke 2003, McKinney 2008, Wolch, Byrne and Newell 2014). Remote sensing has been playing a key role in ISA mapping. To date, the majority of ISA mapping literatures used multispectral sensors, for example MODIS, Landsat ETM+, and QuickBird, etc. Thus, more efforts have been oriented to spatial heterogeneity and much less effort has been devoted to spectral diversity (Weng 2012). It may due to the lack of economic, timely, and global hyperspectral data. The future launch of EnMAP (Environmental mapping and Analysis Program) hyperspectral satellite in 2019 provides new opportunities for ISA mapping (Kaufmann et al. 2006, Guanter et al. 2015). Although hyperspectral sensors have been studied for more than three decades, most mature hyperspectral sensors are mounted on airplanes: FLI and CASI (Gower et al. 1992), AVIRIS (Vane et al. 1993), Hydice (Rickard et al. 1993), HyMap (Cocks et al. 1998), etc. The only three active hyperspectral satellites are NASA's Hyperion (Pearlman et al. 2003), NASA's HICO (Corson et al. 2008), and ESA's CHRIS (Barnsley et al. 2004). However, the CHRIS and HICO are limited to the visible to near-infrared (VNIR) region (400-1400 nm), and Hyperion has a low signal-to-noise ratio, which limits its feature detection capabilities. Therefore, the EnMAP mission is a milestone towards a comprehensive hyperspectral observation from space (Guanter et al. 2015). The designed EnMAP hyperspectral sensor has a 30 m spatial resolution and its spectral range is between 420 and 2450 nm with a spectral sampling distance varying between 5 and 12 nm. The global revisit capability of EnMAP hyperspectral satellite is 21 days. The potential of using EnMAP images in urban environments has been extensively discussed by Heldens et al, 2011, and ISA mapping has been recognized as one of the great strengths of the EnMAP hyperspectral sensor (Heldens et al. 2011).

## 2.1 Background

### 2.1.1    Previous studies on EnMAP

An EnMAP end-to-end simulation tool (EeteS) was designed by Segl et al. (Segl et al. 2012) for calibrating the fundamental instrument parameters, developing the data pre-processing steps, and evaluating the exploitable algorithms. The EeteS software simulates EnMAP images using reflectance data that has spectrally and spatially oversampled resolution than the EnMAP final sampling interval. To date, 18 publications have used the simulated EnMAP images for various earth observation applications (Braun, Weidner and Hinz 2012, Okujeni, van der Linden and Hostert 2015, Rogge et al. 2014, Schwieder et al. 2014, Suess et al. 2015, Yokoya, Chan and Segl 2016, Locherer et al. 2015, Marcinkowska-Ochtyra et al. 2017, Fassnacht, Weinacker and Koch 2011, Lehnert et al. 2014, Dotzler et al. 2015, Steinberg et al. 2016, Clasen et al. 2015, Malec et al. 2015, Leitão et al. 2015, Xi et al. 2015, Siegmann et al. 2015, Mielke et al. 2014) (Table 1), most of which target natural or agricultural environments and only one is in urban environments. In order to produce classification maps or predict certain geographic features, 14 of the publications used the endmembers or training samples obtained from other high-resolution images or existed reference data. Only four publications extracted endmembers directly from simulated EnMAP images, but they are large-scale studies in natural environments, where land cover changes less rapidly than in urban environment. The reason why we want to extract endmembers directly from remote sensing imagery is that field/lab reference data is acquired under different conditions from the airborne or spaceborne data and can cause errors in following analyses.

**Table 2-1: Publications on the application of simulated EnMAP images.**

| Year | Authors | Applications | Endmember extraction methods/ training sample sources | Methods |
|---|---|---|---|---|
| 2017 | Marcinkowska-Ochtyra, et. al. | Per-pixel vegetation classification in subalpine and alpine areas. | Field survey reference map and pixel purity index (PPI) from the airborne APEX images. | Support vector machines (SVM). |
| 2016 | Yokoya, et. al. | Sub-pixel geologic material classification in temperate bare rock area. | Vertex component analysis (VCA) and visual inspection from reference image. | Coupled nonnegative matrix factorization (CNMF) fusion (Yokoya, Yairi and Iwasaki 2012) and multiple endmember SMA (MESMA) (Roberts et al. 1998b). |
| 2016 | Steinberg, et. al. | Predict common surface soil properties | Ground truth measurements. | Partial least squares regression (PLSR) (Oldenburg, Schmidtlein and Feilhauer). |
| 2015 | Okujeni, et. al. | Sub-pixel land cover classification in urban-rural gradient area. | Reference spectral library from airborne HyMap images. | Support vector regression (SVR) (Okujeni et al. 2013). |
| 2015 | Suess, et. al. | Per-/sub-pixel shrub cover abundance classification. | Manually selected from simulated EnMAP images. | Support vector classification (SVC) and adapted SVC (classification). |
| 2015 | Clasen, et. al. | Sub-pixel forest crown classification. | Field survey measurements. | MESMA. |
| 2015 | Malec, et. al. | Sub-pixel soil degradation cover abundance classification. | Extracted from the EnMAP image using updated spatial spectral endmember extraction tool (SSEE) method (Rogge et al. 2012). | MESMA. |
| 2015 | Leitão, et. al. | Predict shrub cover abundance classification. | Reference map. | Boosted regression tree (BRT) (Elith, Leathwick and Hastie 2008). |
| 2015 | Locherer, et. al. | Predict leaf area index (LAI) during agriculture growing season. | Field survey measurements. | Look-up-table based inversion of the PROSAIL model. |
| 2015 | Siegmann, et. al. | Predict LAI in wheat field. | Field survey measurements. | Ehlers fusion and PLSR. |
| 2015 | Xi, et. al. | Predict phytoplankton taxonomic groups | Lab measurements. | Similarity index and hierarchical cluster analysis. |
| 2015 | Dotzler, et. al. | Predict drought stress phenomena in deciduous forest communities. | Reference map. | Analyses of variance (ANOVA), and Tukey's HSD post-hoc tests using drought-sensitive spectral indices. |

| Year | Authors | Applications | Endmember extraction methods/ training sample sources | Methods |
|---|---|---|---|---|
| 2014 | Rogge, et. al. | Sub-pixel geologic material classification in subarctic bare rock area. | Extracted from the EnMAP image using updated SSEE method. | Iterative SMA (ISMA) (Rogge et al. 2006) |
| 2014 | Schwieder, et. al. | Sub-pixel shrub cover abundance classification. | Reference map. | SVR (Karatzoglou, Meyer and Hornik 2005), random forest regression (RF) (Liaw and Wiener 2002), and PLSR (Wehrens and Mevik 2007). |
| 2014 | Lehnert, et. al. | Sub-pixel classification of rangeland degradation in the Tibetan Plateau and predict chlorophyll content. | Field survey measurements. | Linear SMA and PLSR. |
| 2014 | Mielke, et. al. | Predict secondary iron bearing minerals | Field survey measurements. | Iron feature depth (IFD) (Mielke et al. 2014). |
| 2012 | Braun, et. al. | Per-pixel land cover classification in alpine foreland and irrigation agriculture area. | Manually selected from simulated EnMAP images. | SVM, import vector machines (IVM) (Zhu and Hastie 2005), and relevance vector machines (RVM) (Tipping and Faul 2003). |
| 2011 | Faßnacht, et. al. | Predict tree-covered areas. | Reference map. | Extended NDVI (normalized difference vegetation index). |

## 2.1.2    Previous studies on spectral mixture analysis

Pixels in EnMAP images with 30 m spatial resolution are expected to be highly mixed, especially in urban scenes. Spectral unmixing is designed for this problem. The spectral mixture analysis (SMA) is one of the mostly used spectral unmixing method. The SMA process usually consists of two steps: endmember extraction and abundance quantification. The traditional geometry-based endmember extraction methods (e. g. N-Findr, pixel purity index (PPI), and vertex component analysis, etc.) only search for the endmember in the original image and their success heavily depend on the abundant presence of pure pixels (Bioucas-Dias et al. 2012). Thus, the geometry-based endmember extraction methods may encounter difficulty when working with EnMAP imagery of 30m spatial resolution, which have limited endmembers in the original image. The statistical methods are potential alternatives to deal with the lack-of-pure-pixel situation (Ma et al. 2014) and the independent component analysis (ICA) is a widely-used method (Nascimento and Dias 2005). The ICA uses the statistical characteristics of hyperspectral

data to identify the potential endmember. However, a key assumption in ICA that the abundance matrix is statistically independent is compromised due to the sum-to-one and nonnegativity constraint on the abundance matrix in the SMA (Nascimento and Dias 2005).

Another rising group of SMA methods based on nonnegative matrix factorization (NMF) addresses both the problems of lack-pure-pixel and statistical dependency. The NMF has been successfully applied in hyperspectral image SMA over the past 10 years. The NMF is based on the algebraic feature of images (Miao and Qi 2007). Because of the often found nonnegativity in the underlying models, Paatero and Tapper first suggested the positive matrix factorization in 1994 for dimension reduction (Paatero and Tapper 1994). The NMF method factorizes the original data matrix *V* into one basis matrix *W* and one encoding matrix *H*. The fundamental NMF equation (Equation (2-4)) is analogous to the SMA objective function, by reinterpreting base matrix *W* and encoding matrix *H* separately as endmember and abundance matrix. The solution space of the NMF method is in the entire nonnegative space, which is larger than the original image. Thus, the NMF method can find the endmember that do not exist in the original image. Figure 2-1 shows the advantages of NMF method over the geometry-based method. At the left side, in spectral space, the light blue area represents the image simplex. Only one of the three true endmembers E3 exists in the image simplex. As the geometry-based endmember extraction methods can only locate the pixels in the image simplex as endmemebrs, the mixed pixels M1 and M2 may be selected by the geometry-based endmember extraction methods as endmemebrs. At the right side, the solution space of the NMF in light orange color is much larger than the original image space in light blue, which makes it possible to find true endmembers E1 and E2 that are outside of the image simplex.
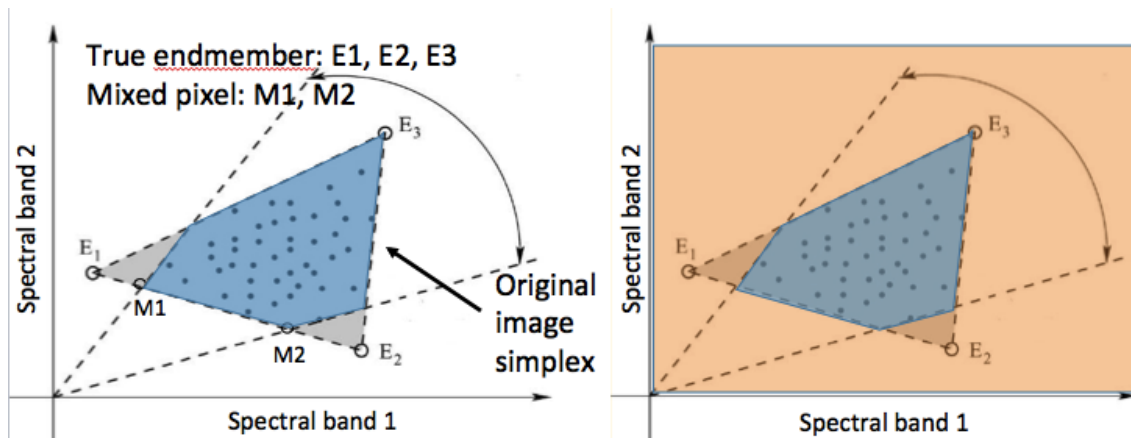
**Figure 2-1: Original image simplex and solution space.**

However, with only the nonnegativity constraint, the solution space of the NMF is very large and prone to inefficient convergence. To address this defect, multiple extensions of the NMF have been proposed: symmetric NMF, semi-NMF, non-smooth NMF, and multilayer NMF (Cichocki et al. 2009). Among different extensions, three are very suitable for the task of SMA, as they accurately depict certain characteristics of the hyperspectral imagery. They are sparse, minimum volume, and nonlinearity extensions. In details, the sparse NMF (sNMF) put the constraint (Bioucas-Dias et al. 2012, Lu et al. 2013) on the abundance matrix accounts for the fact that in one image most of the pixels are mixtures of only a few of the endmembers, instead of using all the reference endmembers . It prefers the sparse abundance matrix ($H$) with a lot of zero entries. For example shown in Figure 2-2, the whole image area has eight different endmembers listed at left. However, each pixel in the image is not likely to have all the eight land covers, but only a few of them. This will lead to a lot of zeros in the abundance matrix.

| | Pixel1 | Pixel2 | P3 | P4 | P5 | P6 | P7 |
|---|---|---|---|---|---|---|---|
| Abundance of endmember 1 | 0 | 1 | 0 | 0 | 0 | 0.3 | 0 |
| Abundance of endmember 2 | 0.5 | 0 | 0 | 0 | 0 | 0 | 0.1 |
| Abundance of endmember 3 | 0 | 0 | 0.6 | 0 | 0.5 | 0 | 0.2 |
| Abundance of endmember 4 | 0.2 | 0 | 0 | 0 | 0.5 | 0.7 | 0.7 |
| Abundance of endmember 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Abundance of endmember 6 | 0.3 | 0 | 0 | 0 | 0 | 0 | 0 |
| Abundance of endmember 7 | 0 | 0 | 0.4 | 0 | 0 | 0 | 0 |
| Abundance of endmember 8 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |

**Figure 2-2: Sparse abundance matrix.**

The minimum volume NMF (vNMF) has the constraint working on the endmember matrix ($W$) through minimization of the simplex volume determined by the endmembers among all possible simplexes that circumscribe the data scatter space (Miao and Qi 2007). This constraint is based on one effective endmember extraction criterion that the vertices of a minimum-volume simplex enclosing all the observed pixels should serve as high-fidelity estimates of the endmembers (Craig 1994). In other words, the selected endmembers should not be too far away from the original image space. In Figure 2-3, the green simplex on the left is circumscribed by the preferred endmembers; while the red simplex on the right though encloses all the image pixels does not serve as high-fidelity estimates of the endmembers, as the vertices are too far away from the original image area.
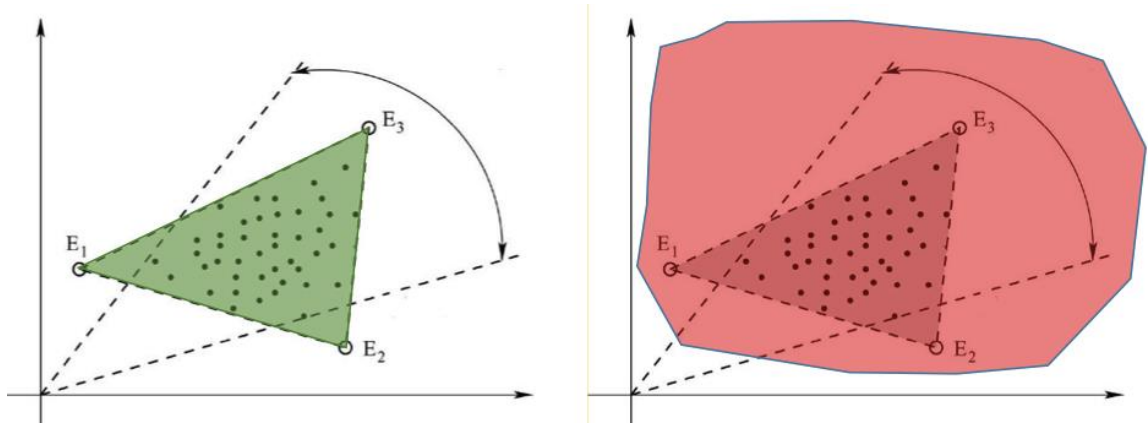
**Figure 2-3: Left: preferred solution; Right: not preferred solution simplex.**

In the end, the traditional NMF is under a linear mixture framework. The linear spectral mixture analysis works well in most of the situations, when the incident radiation is directly reflected back to the sensor. In some complicated situations shown in Figure 2-4, second-order scattering happens as the red line, when the incident radiation first hit the tree, then the soil beneath and in the end back to the sensor. Then, the received reflectance by the sensor is not a linear combination of the existed endmembers in that pixel. The robust nonlinear NMF (rNMF) (Févotte and Dobigeon 2015) introduces an additional term to address possible nonlinear effects.
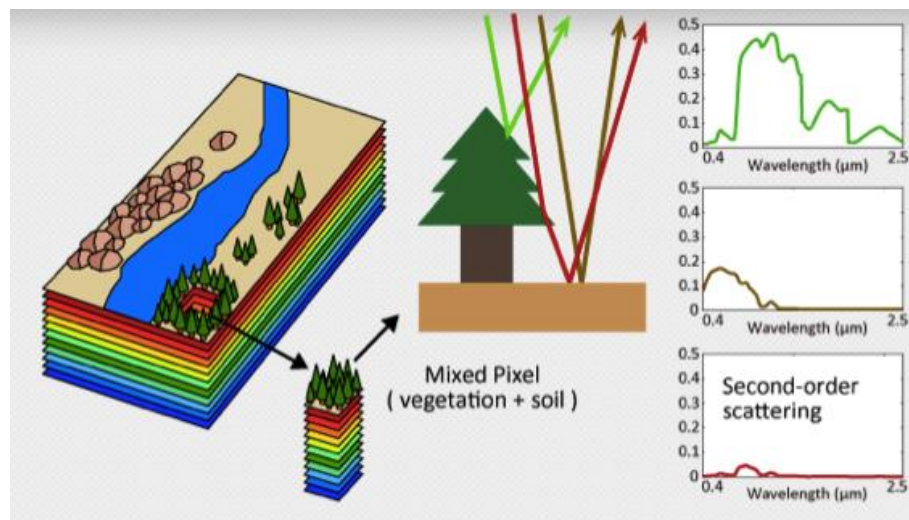


**Figure 2-4: Nonlinear effect (Yokoya, Chanussot and Iwasaki 2014).**

Furthermore, although the NMF-based methods simultaneously obtained the endmembers and abundance matrix, the current NMF studies focus more on the resulting endmembers than the resulting abundance matrices.

## 2.2 Objective

The main objective of Chapter 2 is to find the appropriate image processing methods for ISA mapping using the new hyperspectral satellite EnMAP image. Corresponding to the objective, the following research questions are the focus of Chapter 2:

(1) Among the currently available SMA methods: geometry-, statistic-, and algebra-based methods, which one is suitable for EnMAP ISA mapping?

(2) Compared to the airborne high spatial resolution hyperspectral image, how does the EnMAP perform?

(3) Are the abundance maps directly obtained from the algebra-based methods usable?

(4) What differences can be found between the variations of the algebra-based methods?

(5) How can the research results help the current ISA mapping study?

## 2.3 Methods

### 2.3.1 Hyperspectral subspace identification

For most of the endmember extraction methods, the number of wanted endmembers remains an important input. The decision on this number is part of the problem called hyperspectral subspace identification or virtual dimensionality (VD). This number closely relates to the intrinsic dimensionality of remote sensing imagery. If the set number is too small, the selected endmembers may still be mixtures of several materials, and if the set number is too large, the selected endmembers may be strongly affected by noise (Luo et al. 2013). One popular method that solves this problem is the Harsanyi–Farrand–Chang method (HFC) proposed by Chang and Du in 2004 (Chang and Du 2004). The HFC is

defined by the minimum number of spectrally distinct signal sources, which characterizes the hyperspectral image from the perspective view of target detection and classification (Chang and Du 2004). The difference between the eigenvalues of the correlation matrix $R_{L \times L}$ and covariance matrix $K_{L \times L}$ of the image is used to detect whether a signal source is present in a specific dimension. The Neyman-pearson detection theory is applied in HFC to judge the hypotheses of zero difference ($H_0$) between the mentioned eigenvalues and positive difference ($H_1$) between them. If $H_1$ is true, it means that an endmember is contributing to the correlation eigenvalue in addition to noise. It is because by assuming that signal sources are nonrandom unknown positive constants and noise is white with zero mean, the equations below can be expected (Chang and Du 2004):

$$\hat{\lambda}_l > \lambda_l, \text{ for } l = 1, \dots, VD \text{ and} \tag{2-1}$$

$$\hat{\lambda}_l = \lambda_l, \text{ for } l = VD + 1, \dots, L,$$

where $\hat{\lambda}_l$ is the $l$th eigenvalue of $R_{L \times L}$; and $\lambda_l$ is the $l$th eigenvalue of $\mathrm{K}_{\mathrm{L} \times \mathrm{L}}$. The probability of the hypothesis is made using the false-alarm probability and detection power. The HFC algorithm used in this research is from the Endmember Induction Algorithms toolbox, created by the Computational Intelligence Group of University of the Basque Country (Computational Intelligence Group).

## 2.3.2    Endmember extraction and spectral mixture analysis

The remote sensing imagery with medium spatial resolution, like EnMAP imagery, faces the mixed pixel problem. To solve this problem, spectral unmixing was proposed in the early 1980s (Dozier 1981), and has since been widely adopted, especially in heterogeneous urban environment (Powell et al. 2007, Roberts et al. 1998a, Wu and Murray 2003, Liu et al. 2004). The spectral unmixing method models the reflectance value of the target pixel from more than one endmember. Before performing the spectral unmixing, one needs a group of endmembers as input. This research applied three traditional and four NMF-based endmember extraction methods: N-Findr, PPI (pixel purity index), ICA (independent component analysis), NMF (nonnegative matrix

factorization), sNMF, vNMF, and rNMF. The rest of this section explains each of the methods in detail.

The N-Findr was originally proposed by Winter et al. in 1999. This method automatically finds the extreme points of an n-dimensional scatter plot by finding the simplex of the maximum volume that can be inscribed within the hyperspectral data set (Winter 1999). The initialization of N-Findr is to randomly select pixels as endmembers, which are put into a matrix $E$. A trial volume is calculated for $E$ by the equation (Plaza et al. 2004, Winter 1999):

$$V(E) = \frac{1}{(E-1)!} abs(|E|).$$ (2-2)

The initial volume estimate will be refined through iteration. Every pixel in the image will be sequentially put into matrix $E$ to replace an original endmember. Then, the volume of the new matrix will be calculated. The new pixel replaces the original endmember and stays in the matrix when the replacement results in a volume increase. This procedure is repeated until no replacements of endmembers are left (Winter 1999).

The PPI was initially proposed by Boardman et al. in 1995 (Boardman, Kruse and Green 1995), and this was when the concept of pixel purity was created. The pixel purity value is evaluated in a way that extreme pixels achieve higher scores. Thus, pure pixels have a high pixel purity value and mixed pixels have a low pixel purity value. The PPI is computed by repeatedly projecting n-Dimension scatter plots on a set of random unit vectors called skewers. During the computation, the PPI algorithm records the extreme pixels in each projection (the pixels that fall onto the ends of the unit vector) and notes the total number of times each pixel is marked as extreme. These counts have been associated with pixel purity using a convex geometry argument: a larger count means a purer pixel (Boardman 1993). A pixel purity image is created where each pixel value corresponds to the number of times that pixel was recorded as extreme. The result of PPI algorithm heavily depends on the randomly generated initial set of skewers. Since this initial skewer set is randomly generated, the final results of PPI fails to be consistent and reproducible (Chang, Wu and Chen 2010).

The N-Findr and PPI methods only search endmembers from the original data. When the pure pixel in the image is limited, these two methods may fail. The statistical method ICA provides a solution. The ICA was originally proposed by Hyvärinen and Oja in 1997 (Hyvärinen and Oja 2000). It is a blind source separation method that aims to solve both the mixing matrix $A$ and the source matrix $s$ from the observed matrix $x$:

$$x = As. \tag{2-3}$$

In order to solve this problem, the ICA algorithm applies statistical independency as a criterion on the source matrix. The aim of ICA is to find a linear representation of nongaussian data so that the components are statistically independent, or as independent as possible. Mathematical proof has been given to justify that independency of a variable is equivalent to its nongaussianity (Hyvärinen and Oja 2000). Thus, the ICA algorithm: FastICA adopts a fixed-point iteration scheme to maximize the nongaussianity of the components in source matrix. The fundamental equation (Equation (2-3)) of ICA fits the hyperspectral spectral unmixing problem when taking the mixing matrix as endmember, source matrix as abundance, and observed matrix as original image. Wang and Chang (Wang and Chang 2006) have proposed a complete process of using ICA to simultaneously solve endmember and abundance and achieved great results. However, the sum-to-one and nonnegativity constraints on abundance matrix violate the mutually independent assumption in ICA. As a result, these constraints are added through normalization, which is not a fully constrained approach.

On the contrary, the NMF-based SMA methods naturally embrace the nonnegativity constraint. They also simultaneously calculate the endmember and abundance matrices. Although the cost function of the NMF-based methods is nonconvex, various additional constraints improve the performance of NMF-based methods. The task of the NMF is to find an approximate factorization:

$$V \approx WH, \tag{2-4}$$

where $V$ is the original image with a dimension of l×n; $W$ is the endmember matrix with a dimension of l×k ; $H$ is the abundance matrix with a dimension of k×n ; and $k$ is the

number of wanted endmembers given by HFC. The difference between *V* and *WH* is measured by an adjusted Kullback-Leibler divergence in this paper (Févotte and Dobigeon 2015):

$$D(A||B) = \sum_{ij}(A_{ij}\log\frac{A_{ij}}{B_{ij}} - A_{ij} + B_{ij}). \tag{2-5}$$

The Kullback-Leibler divergence is a non-symmetric difference measure between two matrices, measuring the divergence of matrix *A* from matrix *B* by assessing the information lost when *A* is used to approximate *B*. The NMF algorithm aims to minimize the difference between the original matrix and the reproduced matrix:

$$\text{minimize } f(W, H) = D(V \parallel WH). \tag{2-6}$$

To achieve this goal, the algorithm iteratively updates the endmember and abundance matrices using a multiplicative update rule (Févotte and Dobigeon 2015):

$$H_{a\mu} \leftarrow H_{a\mu} \frac{\sum_i W_{ia} V_{i\mu}/(WH)_{i\mu}}{\sum_k W_{k\alpha}}, \tag{2-7}$$

$$W_{ia} \leftarrow W_{ia} \frac{\sum_\mu H_{a\mu} V_{i\mu}/(WH)_{i\mu}}{\sum_v H_{av}}.$$

This update is proven to keep the divergence non-increasing and converging during the iteration (Zhou, Peng and Chen 2015) and the derivation of the converging rule can be found in the Appendix B. The start *W* and *H* are randomly generated positive matrices and keep positive during the multiplicative update. In the case of SMA, an additional sum-to-one constraint is needed for the abundance matrix *H*. This study adopted the variable substitution approach, which guarantees that the constraint is fulfilled (Arngren, Schmidt and Larsen 2011). The standard NMF process for SMA is shown in the Figure 2-5 below.
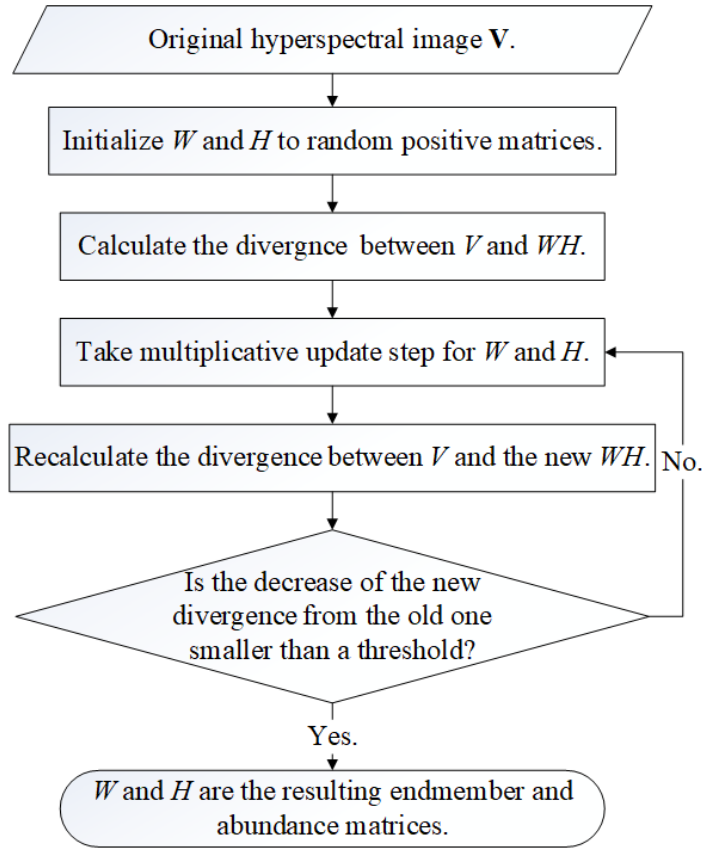
**Figure 2-5: Nonnegative matrix factorization algorithm flowchart.**

In order to include other specific constraints in NMF process, the most common way is to add loss function to the original objective function. The loss functions are usually determined by the regularization of either *W* or *H*. In the case of sNMF, a regularization term of abundance matrix *H* is added to Equation (2-6) and the new objective function becomes:

$$minimize\ f(W, H) = D(V \parallel WH) + \alpha \parallel H \parallel_1, \qquad (2\text{-}8)$$

where $\parallel H \parallel$ is the *L1* regularizer of *H*; and $\alpha$ is the sparseness parameter. The update rule correspondingly changes to:

$$H_{a\mu} \leftarrow H_{a\mu} \frac{\sum_i W_{ia} V_{i\mu}/(WH)_{i\mu}}{(1+a)\sum_k W_{ka}}. \qquad (2\text{-}9)$$

The sparseness parameter $\alpha$ is a key value, as it controls the impact of the sparse

constraint, where a larger value means greater force on the sparse constraint. Besides the fact that this parameter is dependent on the sparseness of the abundance matrix, the value of the sparseness parameter works implicitly in the objective function. Some studies (Qian et al. 2011, Lu et al. 2013) estimate the sparseness parameter according to the sparseness criteria in (Hoyer 2004):

$$\alpha = \frac{1}{\sqrt{L}} \sum_l \frac{\sqrt{N} - \|x_l\|_1 / \|x_l\|_2}{\sqrt{N-1}}, \tag{2-10}$$

where $L$ is the number of bands; $N$ is the number of pixels; $x_l$ is the given hyperspectral data at band l; $\| x_l \|_1$ is the L1 regularizer; and $\| x_l \|_2$ is the L2 regularizer. This study refers to Equation (2-10) to obtain the potential sparseness parameter $\alpha$. In addition, this study tests eight other parameters (0.01, 0.1, 0.2, 0.5, 0.7, 1, 5, and 10) as the sparseness parameters in order to test the sensitivity of sNMF on the parameter.

In the case of vNMF, methods that directly measure volume often use determinant as their regularization on the endmember matrix ($W$) (Miao and Qi 2007, Schachtner et al. 2009). However, when the regularization is strong, the $K$-dimensional volume will collapse to a ($K$-$I$)-dimensional subspace and volume becomes zero, causing a regularization failure. To avoid this defect, a Euclidean distance measure constraint on the endmember matrix ($W$) is proposed (Arngren et al. 2011):

$$minimize \ f(W, H) = D(V \parallel WH) + \beta J_w(W), \tag{2-11}$$

$$J_w^{dist}(W) = \sum_{k=1}^{K} \| w_{:k} - \frac{1}{K} \sum_{k'}^{K} w_{k'} \|_{2.}^2,$$

where $\beta$ is the distance parameter; $\| w_{:k} - \frac{1}{K} \sum_{k'}^{K} w_{k'} \|_{2.}^2$ is the $L^{\wedge}2$ regularizer (Euclidean distance); and $\frac{1}{K} \sum_{k'}^{K} w_{k'}$ denotes the center of the current selected endmembers. The distance parameter controls the impact of the distance constraint. To the best of our knowledge, among current studies concerning the minimum-volume NMF, no illustrations on how to select this parameter can be found, except that the distance parameter is usually between 0 and 1 (Arngren et al. 2011, Miao and Qi 2007). Therefore,

we tried eight distance parameters: 0.01, 0.1, 0.2, 0.5, 0.7, 1, 5, and 10, in order to provide some insights.

Lastly, the rNMF method adds a residual term to the standard NMF factorization Equation (2-4) accounting for the outlier caused by nonlinear spectral effects (Févotte and Dobigeon 2015):

$$V \approx WH + R \,, \tag{2-12}$$

where $R$ is the residual matrix. Then the objective function is changed to:

$$\text{minimize } f(W, H) = D(V \parallel WH + R) + \gamma \parallel R \parallel_{2} ., \tag{2-13}$$

where $\gamma$ is the nonlinearity parameter; $\parallel R \parallel_{2.}$ is the $L^{\wedge}2$ regularizer. The nonlinearity parameter controls the impact of the nonlinearity constraint. Févotte and Dobigeon (Févotte and Dobigeon 2015) proposed a method to estimate the potential nonlinearity parameter value based on results in (Lee et al. 2010):

$$\gamma = \frac{C}{\hat{\mu}}, \tag{2-14}$$

$$C = \frac{2}{\sqrt{\pi}} \cdot \frac{\Gamma(\frac{k}{2}+1)}{\Gamma(\frac{k}{2}+\frac{1}{2})},$$

where $\Gamma$ is Gamma function; $k$ is the number of wanted endmembers given by HFC; $l$ and $n$ are separately the spectral band number and pixel number of the image. This study refers to Equation (2-14) to obtain the potential nonlinearity parameter. However, in (Févotte and Dobigeon 2015), it is specifically pointed out that Equation (2-14) is only a convenient gross estimation without statistical guarantee. Therefore, we tried eight more nonlinearity parameters: 0.01, 0.1, 0.2, 0.5, 0.7, 1, 5, and 10. Using iteration, the endmember, abundance, and residual matrices are sequentially updated. The endmember and abundance matrices update rule are listed in Equation (2-7). The residual update rule is as below (Févotte and Dobigeon 2015):

$$R_{\alpha\mu} \rightarrow R_{\alpha\mu} \frac{\sum_{i} V_{i\mu}/(WH)_{i\mu}}{WH_{i\mu}^{-1} + \lambda \frac{R_{\alpha\mu}}{\parallel R_{\alpha\mu} \parallel_2}}, \tag{2-15}$$

## 2.3.3    Abundance quantification

This study uses the multiple endmember SMA (MESMA) to calculate the abundance matrix for the endmembers derived from N-Findr, PPI, and ICA methods. The MESMA is an extension of the traditional SMA by allowing the number and types of endmembers to vary on a per-pixel basis (Roberts et al. 1998a). It overcomes the limitations of SMA by requiring a model to meet minimum fit and fraction and residual constraints while testing multiple models for each image pixel. MESMA is typically implemented by developing a spectral library, then unmixing an image using every possible combination of two, three and four endmembers applied to each pixel (Roberts et al. 1998a). The VIPER Tools, a plugin for ENVI, developed in Dr. Dar Robert's lab at University of California, Santa Barbara (UCSB) Geography Department provides the MESMA method and is used in this study.

## 2.3.4    Evaluations

This research separately assessed the estimated endmembers and the predicted abundance matrices. We calculated the spectral angle distance (SAD) between each estimated endmember and reference spectra. The SAD is the cosine of the spectral angle formed by n-dimension vectors:

$$SAD = \cos^{-1}(\frac{a^T\hat{a}}{\|a\|\|\hat{a}\|}) \ , \tag{2-16}$$

where $a$ is a reference endmember vector; $\hat{a}$ is an estimated endmember vector. This measurement is invariant before unknown multiplicative scaling that relates to differences in illumination and angular orientation. As the estimated endmember is assigned with the land cover whose reference spectrum has the smallest SAD from the estimated endmember, current literatures often used the minimum SAD to evaluate extracted endmembers: smaller the minimum SAD means better quality of the extracted endmembers. This research found that the estimated endmember with smaller minimum SAD value also has smaller median SAD values with other reference spectra, indicating a smaller SAD with other reference spectra. A good endmember of significant spectral

signature should be close to the target reference spectrum and be far from other reference spectra. Thus, we also considered the median SAD in our evaluation.

To evaluate the abundance matrix, the research used two commonly used indexes: the Pearson correlation coefficient of the linear regression model between the reference and predicted ISA percentages, and the mean absolute difference (MAD) between the reference and predicted ISA percentages (Wang and Chang 2006, Hoyer 2004). In addition, the research proposed a new abundance matrix evaluation method: the overall accuracy of the dominant abundance classification. The NMF-derived abundance matrices were aggregated into ISA and pervious surface area abundance maps and compared with a reference ISA abundance map. For the simulated EnMAP image, the reference ISA abundance map was derived from the airborne CASI images. The airborne CASI images divided the studied area into 144 tiles, and each tile covers about $1.6km \times 1.6km$. We grouped the 144 tiles into 18 mosaic images corresponding to the 18 flight lines in order to lighten the workload of classifying individual tiles. In the same flight line, the condition during image capture is assumed to be consistent. The 18 mosaic images were first classified into 7 to 9 classes using the maximum likelihood, minimum distance, and spectral angle mapper per-pixel classification methods, among which the maximum likelihood method provides the best accuracy. The initial classification maps then aggregated into impervious/pervious surfaces. Five among the 18 mosaic reference classification maps were visually evaluated using 500 sample points for accuracy assessment. The classification overall accuracy range from 85.5% to 89.4%. The reference impervious and pervious surface abundance maps were calculated within squares of $30 \times 30$ CASI pixels, as each EnMAP pixel (30 m resolution) overlays with 900 CASI pixels (1 m resolution) (Figure 2-6). For the Hydice image, the reference abundance matrix is available online (Zhu et al. 2014). For both of the reference impervious surface maps, the number of both pixels with larger than 99% and smaller than 1% imperviousness that can be considered as pure pixels are lower than 20%. Thus, the majority of the pixels are mixed pixels.
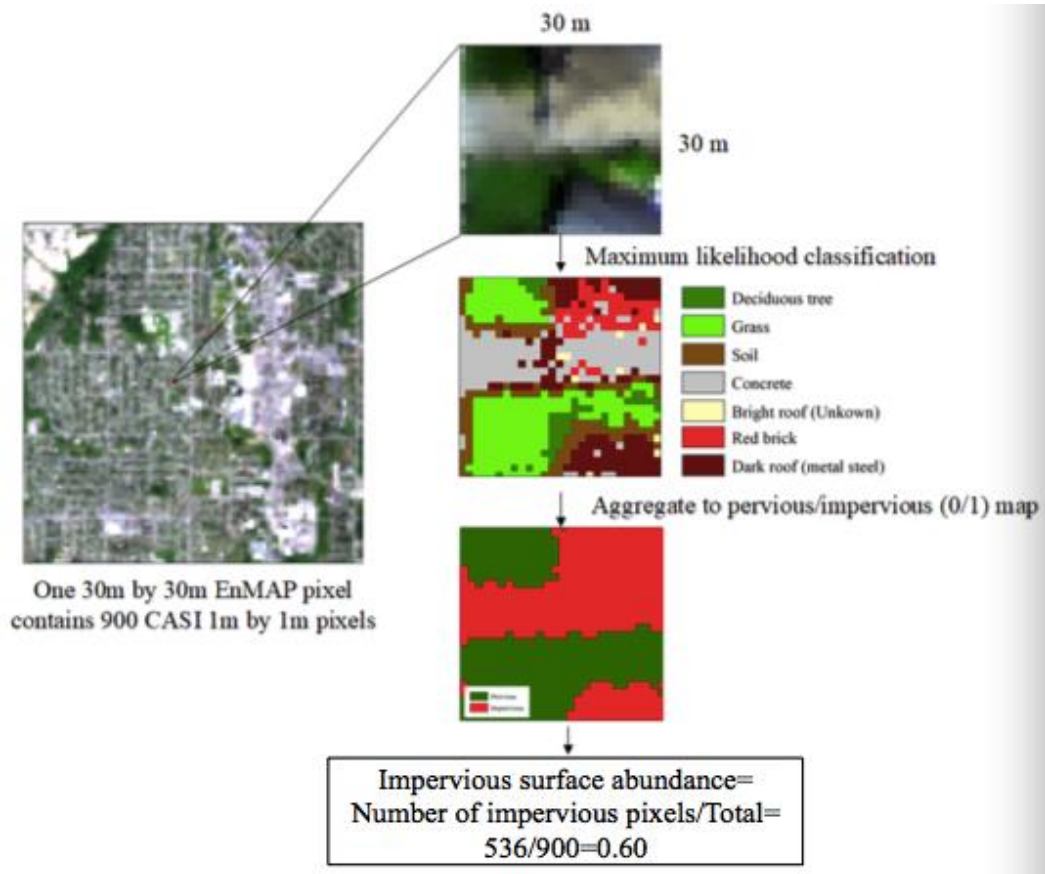
**Figure 2-6: Generate reference map from the CASI image.**

A linear regression model is fitted to a sample set from the reference and predicted ISA abundance maps. The Pearson correlation coefficient $R$ is used to evaluate the strength and direction of a linear relationship. It ranges from -1 to +1, where -1 suggests the strongest possible disagreement and +1 suggests the strongest possible agreement between the two variables. It is the abundance of the variation in dependent variable that is explained by the regression. For two variables $X$ and $Y$, their correlation coefficient $R$ can be derived from the equation below:

$$R = \frac{cov(X,Y)}{\sigma_X \sigma_Y},$$ 

(2-17)

where $cov$ is the covariance; $\sigma_X$ is the standard deviation of $X$; and $\sigma_Y$ is the standard deviation of $Y$. To test for reliability of the observed correlation, the P-value is calculated

for each Pearson correlation coefficients. The P-value is the probability that the observed correlation is in fact zero, also referred to as null hypothesis. It is done by the analysis of the frequency distribution of target value, in my case the correlation coefficient $R$. Normally, a P-value that equals to or is smaller than 0.05 is considered acceptable, as it indicates a less than 5% possibility of the null hypothesis.

The correlation coefficient $R$ reflects the correlation between two variables, while the MAD quantifying the predicting performance. It measures the mean absolute difference between two variables. For two variables $X$ and $Y$ each having $n$ observations, their MAD can be derived using equation below:

$$MAD = \frac{\sum_{i=0}^{n} x_i - y_i}{n}.$$  (2-18)

Many practical applications require ISA classification maps in addition to ISA abundance maps. Traditionally, classification problems often result in a predicted probability surface, which is then translated into a presence–absence classification map. Here, we replace the probability surface with ISA abundance maps. Then, this translation requires a choice of threshold above which the variable of interest is predicted to be present. The common threshold is 0.5. The results are 0/1 maps, where 0 means pervious surface and 1 means impervious surface. The overall accuracy was calculated as shown below:

$$OA = \frac{N_a}{N},$$  (2-19)

where $N_a$ is the number of correctly classified pixel; and $N$ is the number of the total pixels in the image.

## 2.4 Experiments and Results

### 2.4.1 Study images

The simulated EnMAP imagery used in this research was generated by Segl, using the EeteS (Segl et al. 2012). The original data that has been used to simulate EnMAP was obtained by the airborne sensor CASI-1500. The airborne CASI-1500 captured high

spatial resolution hyperspectral images in Surrey, BC, Canada, during April 2013. It contains 72 bands in VNIR portion (360nm to 1050nm) with a 9.6nm band interval. The spatial resolution of the imagery is 1m. The simulated EnMAP imagery has a 30m spatial resolution and 88 spectral bands covering from 420nm to 990nm. The waterbody in the study area was masked out using the waterbodies region map downloaded from the Scholars GeoPortal, a project of the Ontario Council of University Library. The final studied image is in a size of 806 by 585. After matricization and deletion of the no data pixel, there are 350555 pixels left for processing. The simulated EnMAP imagery is shown in Figure 2-7 with RGB combination.
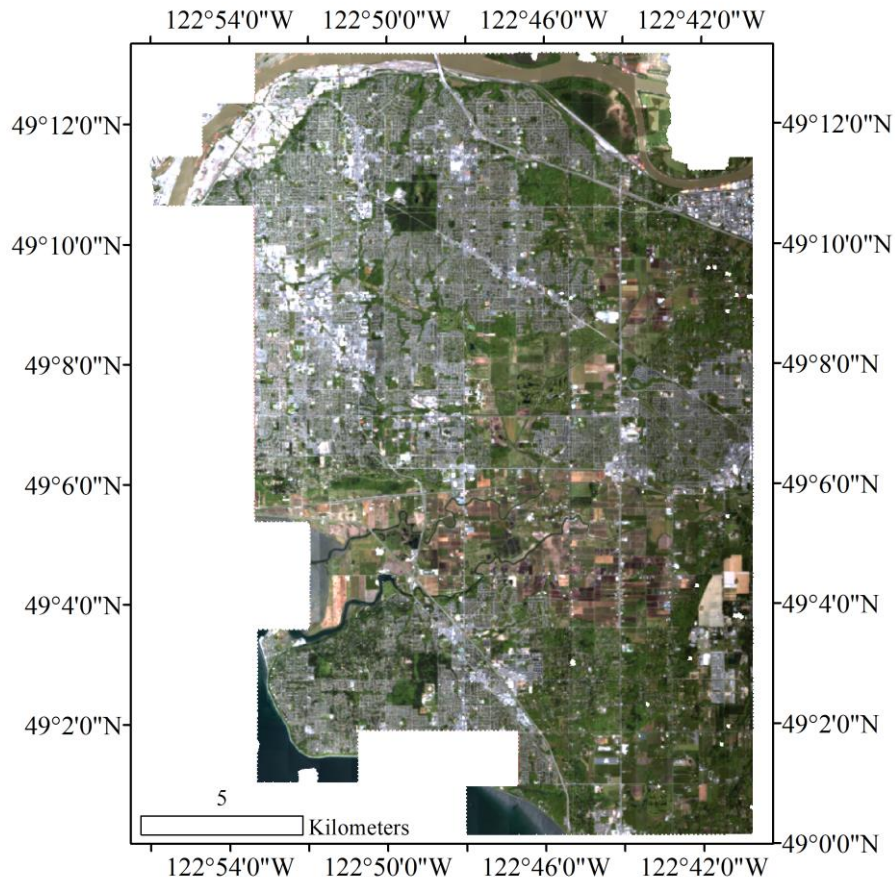


**Figure 2-7: The simulated EnMAP image with RGB combination (Surrey, BC, Canada).**

A real hyperspectral data is also included in this study: the Hydice urban test image. It was recorded by the Hydice sensor in October 1995, and is located in an urban area at Copperas Cove, TX, USA. This image has been widely used in hyperspectral spectral unmixing studies (Zhu et al. 2014). It has a 2m spatial resolution. This image contains $307 \times 307$ pixels and 210 wavelengths ranging from 400nm to 2500nm, resulting in a spectral resolution of 10nm. Due to dense water vapor and atmospheric effects, the channels 1-4, 76, 87, 101-111, 136-153 and 198-210 are often removed. The remaining 162 channels are used in this research. The Hydice urban image is shown in Figure 2-8.



**Figure 2-8: Hydice urban image with RGB combination (Copperas Cove, TX, USA).**

## 2.4.2　　Reference spectra

The potential reference spectra for the simulated EnMAP image were carefully selected from the original CASI images through visual inspection and then further refined using three indices: endmember average RMSE (EAR), minimum average spectral angle (MASA), count based endmember selection (CoB). The basic idea of EAR is to model each pixel within one landcover class using every pixel in the same class and calculate

the endmember average RMSE (EAR) for each endmember. It can be expressed by the equation below (Dennison, Halligan and Roberts 2004):

$$EAR_{A_i,A} = \frac{\sum_{j=1}^{n} RMSE_{A_i,A_j}}{n-1} \qquad (2\text{-}20)$$

where A is one landcover class; n is the number of the endmembers in the potential endmember set from PPI; $A_i$ is $i^{th}$ endmember in this endmember set; $RMSE_{A_i,A_j}$ is the RMSE when using endmember $A_j$ to model endmember $A_i$. Small EAR values mean the endmember is representative for that class, while large EAR value means that the pixel may be an outlier. The cut off threshold of EAR is an experienced value that can be derived from test.

CoB is first proposed in 2003 and it is also based on the result of modelling endmembers using other potential endmembers (Roberts et al. 2003). With the result of modelling, the total number of spectra modeled within the class (in_CoB) and the total number of models outside of the class (out_CoB) when meet the set constraints are recorded for each model. The constraints include the fraction in equation (2) and RMSE. The optimum model is selected as the one that has the highest in_CoB value.

MASA within a class is calculated as the average spectral angle between the reference spectrum and all other spectra within the same class. The best MASA candidate is selected as the one that produces the lowest average spectral angle.

In total, 20 endmembers of 18 different land cover types were visually identified in the CASI images (Figure 2-9). They were resampled using the wavelength and FWHM (full width at half maximum) parameters of the EnMAP sensor. Since ISA is the main goal of this study, these land cover types were grouped according to the V-I-S and the extended V-I-S model (Table 2-2). The V-I-S model was proposed by Ridd (Ridd 1995) to address the challenge of the dramatic amount of heterogeneous materials in the urban environment. The extended V-I-S model was proposed by Okujeni, et al. (Okujeni et al. 2015), especially for EnMAP data, as they found that the EnMAP data is able to map not only the basic V-I-S model but also the extended V-I-S model. The extended V-I-S

model differentiates vegetation into low vegetation and tree, and impervious surface into roof and pavement. In our research, the estimated endmembers from EnMAP were found to achieve best separability when representing more detailed land cover types, shown in the 'Further extended V-I-S model' of Table 2-2.
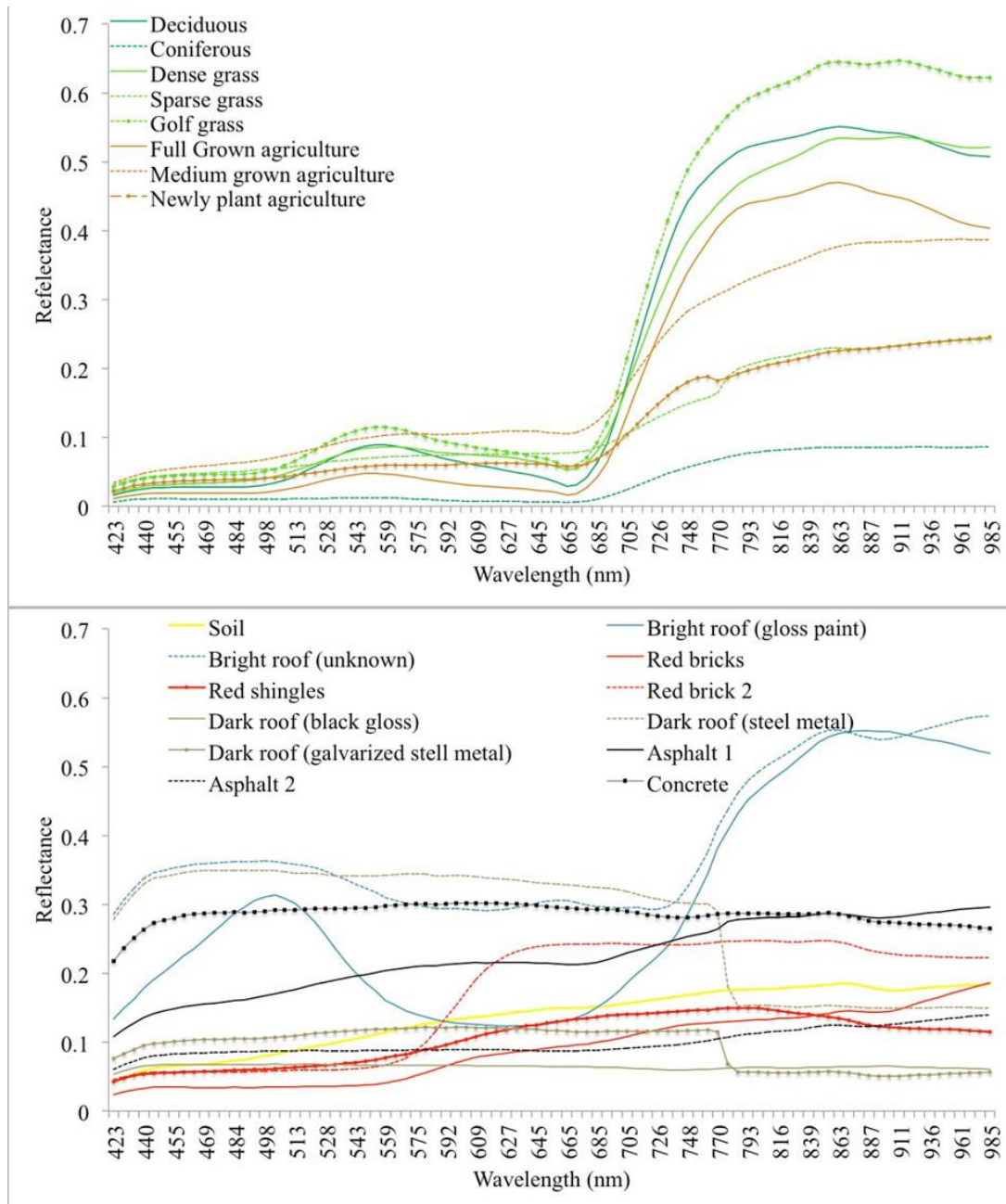


**Figure 2-9: Resampled reference endmembers manually selected from CASI images. (upper: vegetation endmembers; bottom: soil and ISA endmembers)**

**Table 2-2: Reference spectra.**

| V-I-S | Extended V-I-S | Further extended V-I-S | Urban materials | No. Spectra |
|---|---|---|---|---|
| Vegetation | High vegetation | High vegetation | Deciduous tree | 1 |
| | | | Coniferous tree | 1 |
| | Low vegetation | Manicured grass | Intensively manicured grass | 1 |
| | | | Extensively manicured grass | 1 |
| | | | Sparse/dry grass | 1 |
| | | Planted agricultural land | Newly planted agriculture land | 1 |
| | | | Medium grown agriculture land | 1 |
| | | | Fully grown agriculture land | 1 |
| Impervious | Roof | White roof | Gloss paint | 1 |
| | | | Unknown | 1 |
| | | Red roof | Red bricks | 2 |
| | | | Red shingle | 1 |
| | | Dark roof | Black gloss | 1 |
| | | | Galvanized steel | 1 |
| | | | Steel metal | 1 |
| | Pavement | Asphalt | Asphalt | 2 |
| | | Concrete | Concrete | 1 |
| Soil | Soil | Soil | Soil | 1 |

The reference spectra for the Hydice image are available online (Zhu et al. 2014) and are shown in Figure 2-10.
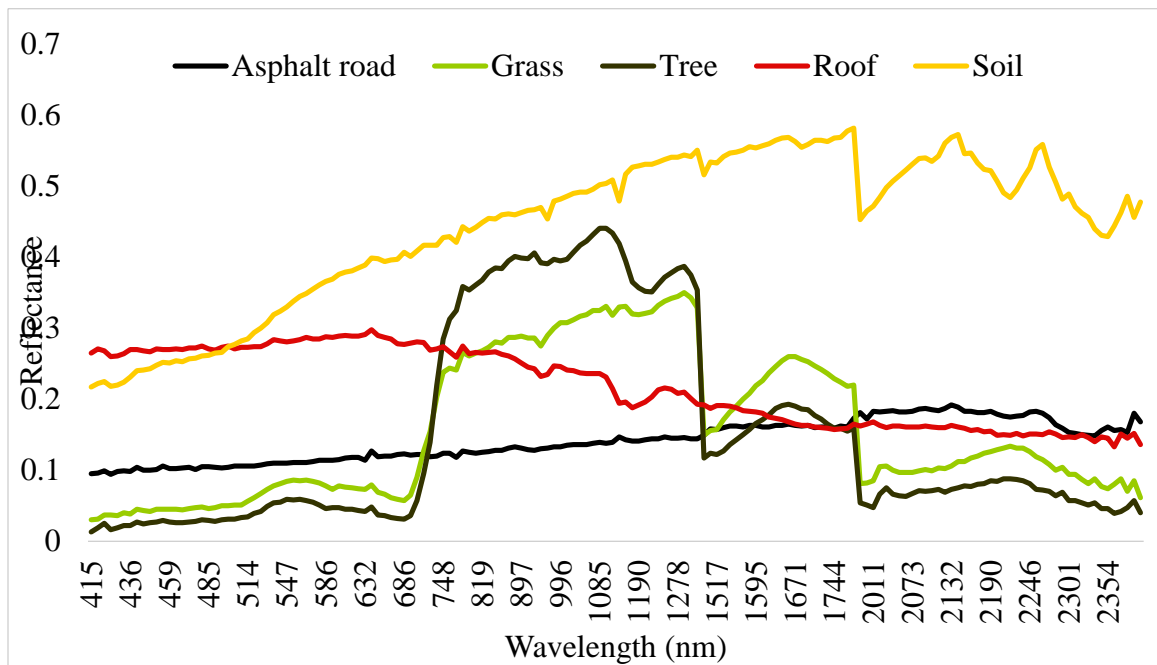


**Figure 2-10: Reference endmembers for Hydice urban image.**

## 2.4.3 Extracted endmember

The subspace of the simulated EnMAP image was found to be 27 by the HFC method, which is used as the target dimension in the SMA. The reason we did not choose to use the number 18 (the number of visually interpreted land cover types) is because the visual interpretation is based on CASI images not the simulated EnMAP image. For the simulated EnMAP image, the sparseness parameter determined by Equation (2-11) was 0.9 and the nonlinearity parameter determined by Equation (2-15) was 0.003. In total, the N-Findr, PPI, ICA, NMF, sNMF, vNMF, and rNMF SMA methods with different constraint parameters on the added constraints composed 30 experiments on the simulated EnMAP image: N-Findr, PPI, ICA, NMF, sNMF(0.9), sNMF (0.01), sNMF (0.2), sNMF (0.5), sNMF (0.7), sNMF (1), sNMF (5), sNMF (10), vNMF(0.01), vNMF (0.1), vNMF (0.2), vNMF (0.5), vNMF (0.7), vNMF (1), vNMF (5), vNMF (10), rNMF (0.003), rNMF (0.01), rNMF (0.1), rNMF (0.2), rNMF (0.5), rNMF (0.7), rNMF (1), rNMF (5), and rNMF (10), where the value in the parenthesis is the used constraint parameter. The derived 30 sets of estimated endmembers were separately compared with the 20 resampled reference spectra, resulting in 30 SAD (spectral angle distance) matrices. For the reference spectra that have similar spectral shapes (e.g. reference soil spectrum and several impervious reference spectra: red roof, asphalt and concrete), it was risky to assign the land cover types based on one SAD. Thus, this research grouped reference spectra based on their land cover types and used the average SAD within each group to assign a land cover group type to the estimated endmembers. We first tried to group the reference spectra according to the V-I-S and the extended V-I-S models (Table 2-2). Both of these models failed to provide satisfactory results. Specifically, the V-I-S model and the extended V-I-S model grouped a wide range of impervious types together and averaged their SAD. Because different ISAs have very different spectral signatures, by averaging their SADs, their particular small SAD may be lost. In the results, the land cover type assignments tended to incorrectly assign soil to impervious endmembers. Therefore, we further partitioned the extended V-I-S model, by separating the roofs to bright, red, and dark roofs, and separating the pavement to asphalt and concrete ('Further extended V-I-S model' column in Table 2-2). In the results, we found that the soil and

ISA can be better separated. We averaged the 30 SAD matrices according to the 9 categories. The land cover types of the EnMAP estimated endmembers were assigned to the category with the smallest average SAD value.

In the case of the Hydice urban image, the HFC method calculated 5 as the subspace number, which was practicable as the reference data provided a 5-endmember (asphalt road, grass, tree, roof, and soil) ground truth set (Figure 2-6). For the Hydice urban image, the sparseness parameter determined by Equation (2-11) is 1.9 and the nonlinearity parameter determined by Equation (2-15) is 0.01. Therefore, the N-Findr, PPI, ICA, NMF, sNMF, vNMF, and rNMF SMA methods with different constraint parameters on the added constraints composed 29 experiments on the Hydice urban image: N-Findr, PPI, ICA, NMF, sNMF(1.9), sNMF (0.01), sNMF (0.2), sNMF (0.5), sNMF (0.7), sNMF (1), sNMF (5), sNMF (10), vNMF(0.01), vNMF (0.1), vNMF (0.2), vNMF (0.5), vNMF (0.7), vNMF (1), vNMF (5), vNMF (10), rNMF (0.01), rNMF (0.1), rNMF (0.2), rNMF (0.5), rNMF (0.7), rNMF (1), rNMF (5), and rNMF (10), where the value in the parenthesis is the used constraint parameter. The derived 29 sets of estimated endmembers were separately compared with the five reference spectra. In the resulting SAD matrices, we found that the single SAD value between each estimated endmember and each reference endmember can successfully identify the land cover types, without any regroup process. The ICA and rNMF (0.2) experiments did not find any impervious endmembers. Thus this experiment cannot further produce any ISA abundance maps or ISA classification maps.

The endmember SAD results are in Table 2-3 for the simulated EnMAP image and in Table 2-4 for the Hydice image. The minimum and median SAD values and their difference within each SAD matrix are shown. As mentioned before, good endmembers with significant spectral signatures should be close to the target reference spectrum and far from other reference spectra, which means smaller minimum values, larger median values, and larger differences between the two. For the constrained NMF methods, only the SAD results with the maximum differences between the minimum and median SAD values are shown. It can be noticed that in both test images the NMF-based method

achieved better endmembers with larger differences SAD values and the constrained NMF performed better than the traditional NMF.

**Table 2-3: Average minimum/median/difference spectral angle distance values (EnMAP).**

| Average SAD | N-Findr | PPI | ICA | NMF | sNMF (0.01) | vNMF (5) | rNMF (5) |
|---|---|---|---|---|---|---|---|
| Minimum | 0.11 | 0.13 | 0.10 | 0.19 | 0.12 | 0.13 | 0.13 |
| Median | 0.45 | 0.45 | 0.41 | 0.55 | 0.55 | 0.57 | 0.56 |
| Difference | 0.34 | 0.33 | 0.31 | 0.36 | 0.43 | 0.44 | 0.43 |

**Table 2-4: Average minimum/median/difference spectral angle distance values (Hydice).**

| Average SAD | N-Findr | PPI | ICA | NMF | sNMF (0.1) | vNMF (10) | rNMF (10) |
|---|---|---|---|---|---|---|---|
| Minimum | 0.29 | 0.30 | 0.18 | 0.12 | 0.16 | 0.13 | 0.16 |
| Median | 0.74 | 0.78 | 0.59 | 0.57 | 0.67 | 0.56 | 0.68 |
| Difference | 0.45 | 0.48 | 0.41 | 0.45 | 0.51 | 0.43 | 0.52 |

## 2.4.4    ISA abundance and classification maps

With the estimated endmembers identified, we aggregated the corresponding abundance matrices to generate ISA abundance maps. For the simulated EnMAP image, the reference impervious abundance map was derived from the original CASI images. For the Hydice urban image, the reference abundance map was available online (Zhu et al. 2014). Then, we randomly selected 500 pair-samples from the predicted and reference ISA abundance maps. In order to evaluate the predicted ISA abundance maps, linear regression models were fitted to the scatter plot of reference and predicted ISA abundance. The correlation coefficients of the fitted linear regression model are listed in Table 2-5 (EnMAP) and 2-6 (Hydice), where the value in the parenthesis after the SMA method is the used constraint parameter. In the original results derived from NMF-based methods, the predicted ISA abundance maps greatly underestimated the ISA abundance, when the linear regression slopes are around 0.5, much less than 1. Thus, normalizations have been performed.

The above linear regression results for the predicted ISA abundance maps show whether the predicted abundance maps closely correlate to the reference abundance map. However, for a lot of practical applications, ISA classification maps are more useful than ISA abundance maps. Thus, we converted the predicted ISA abundance maps to ISA binary classification maps. The task of translating ISA abundance maps to binary pervious/impervious classification maps was similar to the task of converting probability maps to binary classification maps that are commonly applied in probability based classification methods (e.g. maximum likelihood classification). The value 0.5 often serves as a universal threshold. Although there are many adaptive threshold-selecting methods, they were out of the scope of this study. Thus, we used 0.5 as the threshold in our study. The last columns in Table 2-5 and 2-6 show the overall accuracy for the classification results.

For the simulated EnMAP image, the correlation coefficients were less than 0.20 when using the N-Findr, PPI, and ICA methods, which suggested a very bad correlation between the predicted and reference ISA percentages. On the contrary, when using the NMF-based methods, the correlation coefficients ranged from 0.66 to 0.73. Considering the 27 processed NMF-based experiments, the traditional NMF method achieved a moderate correlation coefficient of 0.71. The different selection of the constraint parameters (10, 1, and 10), which are respectively the sparseness, volume, and robust constraints, increased the correlation coefficient by at least 0.02. The sNMF, vNMF, and rNMF methods achieved limited improvement in ISA results than the traditional NMF method. We also found that the parameter values derived from Equations (2-11) and (2-15) for sparseness and nonlinearity constraints, respectively, did not provide the highest linear regression correlation coefficient. The P-values for all the regression models passed the 5% null hypothesis. Regarding the MAD values, the N-Findr, PPI, and ICA methods resulted in more than 0.4 MAD values, suggesting more than 40% difference between the reference and the predicted ISA percentages. The traditional NMF method achieved a 0.1447 MAD value. The constrained NMF methods provided better MAD performances than the traditional NMF when using appropriate constraint parameters. The N-Findr, PPI, and ICA SMA methods resulted in very poor ISA classification

accuracy (~50%). The NMF-based SMA methods easily achieved ~85% overall accuracy. The constrained NMF-based methods did not help to improve the ISA classification results compared to the traditional NMF method.

**Table 2-5: Reference and predicted ISA abundance linear regression parameters and classification overall accuracy (EnMAP).**

| Method | R | P-value | MAD | Accuracy |
|---|---|---|---|---|
| NFindr | 0.20 | 0.85 | 0.4255 | 42.9% |
| PPI | 0.14 | 0.72 | 0.4882 | 57.0% |
| ICA | 0.00 | 0.89 | 0.4597 | 52.9% |
| **NMF** | **0.71** | **4.5e-9** | **0.1447** | **86.7%** |
| sNMF (0.9) | 0.71 | 6.6e-8 | **0.0935** | 86.0% |
| sNMF (0.01) | 0.71 | 5.4e-8 | 0.1513 | 86.0% |
| sNMF (0.1) | 0.69 | 9.0e-8 | 0.1533 | 86.2% |
| sNMF (0.2) | 0.72 | **4.9e-9** | **0.1465** | **86.5%** |
| sNMF (0.5) | 0.67 | 5.0e-8 | 0.1526 | 84.3% |
| sNMF (0.7) | 0.67 | 6.7e-8 | 0.1504 | 85.4% |
| sNMF (1.0) | 0.66 | 5.1e-7 | 0.1578 | 84.7% |
| sNMF (5.0) | 0.70 | 5.8e-8 | 0.1551 | 85.9% |
| **sNMF (10.0)** | **0.73** | 2.3e-8 | 0.1480 | 86.4% |
| vNMF (0.01) | 0.71 | 7.3e-7 | 0.1513 | 86.0% |
| vNMF (0.1) | 0.71 | 5.9e-8 | 0.1468 | 86.1% |
| vNMF (0.2) | 0.70 | 4.6e-8 | 0.1533 | 86.2% |
| vNMF (0.5) | 0.71 | 3.8e-7 | 0.1481 | 85.8% |
| vNMF (0.7) | 0.71 | 9.9e-8 | **0.1464** | 86.6% |
| **vNMF (1.0)** | **0.73** | **1.2e-9** | 0.1473 | 86.4% |
| vNMF (5.0) | 0.71 | 4.5e-7 | 0.1499 | 86.5% |
| vNMF (10.0) | 0.73 | 6.9e-8 | 0.1490 | **86.9%** |
| rNMF (0.003) | 0.69 | 5.5e-8 | 0.1562 | 85.7% |
| rNMF (0.01) | 0.69 | 4.8e-7 | 0.1545 | 85.7% |
| rNMF (0.1) | 0.69 | 3.7e-7 | 0.1533 | 86.2% |
| rNMF (0.2) | 0.71 | 4.2e-7 | 0.1568 | 79.1% |
| rNMF (0.5) | 0.68 | 6.3e-8 | 0.1569 | 85.6% |
| rNMF (0.7) | 0.70 | 7.8e-7 | 0.1530 | 86.4% |

| | | | | |
|---|---|---|---|---|
| rNMF (1.0) | 0.69 | 5.5e-8 | 0.1526 | 85.1% |
| rNMF (5.0) | 0.71 | 4.5e-7 | 0.1510 | 86.1% |
| **rNMF (10.0)** | **0.74** | **1.7e-9** | **0.1500** | **86.7%** |

For the Hydice urban image, the results from the N-Findr and PPI methods are slightly higher, compared to the results for the simulated EnMAP image. This may be because the Hydice urban image has a much higher spatial resolution (2m) and contains less mixed pixels than the simulated EnMAP image. But their correlation coefficients were still low (<0.60). The correlation coefficients from NMF-based methods had a much larger range (0.32 to 0.83) than in the simulated EnMAP case. The traditional NMF methods achieved a moderate linear regression correlation coefficient of 0.70. With the sparseness, volume, and nonlinearity constraints, the highest achieved correlation coefficients were 0.83, 0.80, and 0.70, respectively. The constraint parameter values used for the highest linear regression accuracies are 1 for sparseness, 0.1 for volume, and 5/10 for nonlinearity, which were also not the values provided by the Equations (2-11) and (2-15). The MAD values achieved by the N-Findr and PPI methods were larger than 0.35. The traditional NMF method achieved a 0.1666 MAD value. Two of the constrained NMF methods can improved the MAD values correspondingly to 0.1174 (sparseness) and 0.1325 (minimum volume). The N-Findr, PPI, and ICA methods performed a little better in the Hydice urban image than in the simulated EnMAP image, with over 75% accuracy from the N-Findr and PPI methods. The NMF-based SMA methods also achieved ~85% overall accuracy. The constrained NMF-based methods greatly helped to improve the ISA classification results compared to the traditional NMF method.

**Table 2-6: Reference and predicted ISA abundance linear regression parameters and classification overall accuracy (Hydice).**

| Method | R | P-value | MAD | Accuracy |
|---|---|---|---|---|
| NFindr | 0.57 | 0.56 | 0.4679 | 80.0% |
| PPI | 0.40 | 0.48 | 0.3511 | 75.7% |
| ICA | - | - | - | - |
| **NMF** | **0.70** | **3.6e-9** | **0.1666** | **87.0%** |
| sNMF (1.9) | 0.76 | 6.5e-8 | 0.1463 | 87.2% |
| sNMF (0.01) | 0.71 | 6.3e-8 | 0.1789 | 87.5% |
| sNMF (0.1) | 0.71 | 7.8e-6 | 0.1718 | 85.6% |
| sNMF (0.2) | 0.61 | 8.8e-7 | 0.2051 | 81.6% |
| sNMF (0.5) | 0.69 | 5.7e-9 | 0.1758 | 85.4% |
| sNMF (0.7) | 0.62 | 7.3e-8 | 0.1849 | 83.0% |
| **sNMF (1.0)** | **0.83** | **5.6e-12** | **0.1174** | **91.8%** |
| sNMF (5.0) | 0.46 | 9.7e-9 | 0.2756 | 75.5% |
| sNMF (10.0) | 0.64 | 2.2e-7 | 0.2151 | 79.7% |
| vNMF (0.01) | 0.42 | 3.2e-8 | 0.2542 | 75.9% |
| **vNMF (0.1)** | **0.80** | **5.6e-10** | **0.1325** | **87.9%** |
| vNMF (0.2) | 0.60 | 7.6e-9 | 0.2083 | 81.0% |
| vNMF (0.5) | 0.32 | 3.4e-7 | 0.3139 | 72.8% |
| vNMF (0.7) | 0.71 | 2.7e-8 | 0.1659 | 82.0% |
| vNMF (1.0) | 0.72 | 3.9e-8 | 0.1337 | 85.5% |
| vNMF (5.0) | 0.69 | 1.5e-7 | 0.1778 | 85.8% |
| vNMF (10.0) | 0.71 | 4.6e-8 | 0.1704 | 85.5% |
| rNMF (0.01) | 0.62 | 5.4e-8 | 0.1849 | 83.0% |
| rNMF (0.1) | 0.42 | 7.4e-8 | 0.2394 | 77.6% |
| rNMF (0.2) | - | - | - | - |
| rNMF (0.5) | 0.48 | 9.5e-8 | 0.2291 | 78.6% |
| rNMF (0.7) | 0.61 | 1.3e-7 | 0.2051 | 81.6% |
| rNMF (1.0) | 0.70 | **4.5e-9** | 0.1666 | 87.0% |
| **rNMF (5.0)** | **0.71** | 1.3e-8 | **0.1789** | **87.5%** |
| rNMF (10.0) | 0.71 | 3.7e-8 | 0.1718 | 85.6% |

## 2.5 Discussion

Using N-Findr, PPI, ICA, and four NMF-based (NMF, sNMF, vNMF, and rNMF) SMA methods, 30 sets and 29 sets of ISA abundance and classification maps were generated separately for the simulated EnMAP and Hydice urban images. The results prove great potential for the future EnMAP satellite to handle heterogeneous urban environments. Specifically, what we learned from these results can be discussed in two parts: (1) the comparison of the ISA abundance maps among different SMA methods and between the two study images; (2) the comparison of the ISA classification maps among different SMA methods and between the two studied images.

### 2.5.1    ISA abundance maps

With the estimated endmember sets identified, the corresponding abundance maps were aggregated to obtain the ISA abundance maps. The resulting ISA abundance maps are shown in Figure 2-11 for the simulated EnMAP image and in Figure 2-12 for the Hydice urban image. Only three constrained NMF ISA abundance maps were shown for each image, corresponding to the sNMF, vNMF, and rNMF results with the best correlation coefficients. In the first two rows of Figure 2-11, the reference ISA abundance map is shown, followed by N-Findr, PPI, and ICA. The N-Findr, PPI, and ICA results are not picking up any correct ISA distribution patterns. As the geometry-based N-Findr and PPI methods heavily depend on the presence of pure pixels in the image, the poor performances with the highly mixed image are expected. In the case of ICA, because the independency assumption on the data is violated, the accuracy of ICA-derived abundance values varies greatly between different independent components (endmembers), leading to an inaccurate ISA abundance map. In the third and fourth rows of Figure 2-11, the traditional and the constrained NMF methods generate much better ISA abundance maps than the geometry- and statistic-based methods; they successfully capture the ISA abundance distribution. It can be noticed that the traditional NMF results in lower ISA abundance than the reference ISA abundance map over the entire study area. In circle A, the very high ISA abundance values in the urban built-up areas were overlooked by the

traditional NMF method. In circle B, the medium ISA abundance values in the suburban areas were also overlooked by the traditional NMF method. This underestimation problem was not found in constrained NMF methods. The constrained NMF methods recognized the very high ISA abundance values in circle A. In circle B, both the sNMF and vNMF overestimated the ISA abundance and the vNMF suffered more from an overestimation than the sNMF. Other evidence of this is found in circle C, where the sNMF and vNMF also have the overestimation problem. The rNMF provided smaller ISA abundance values in circle B and C than the sNMF and vNMF.
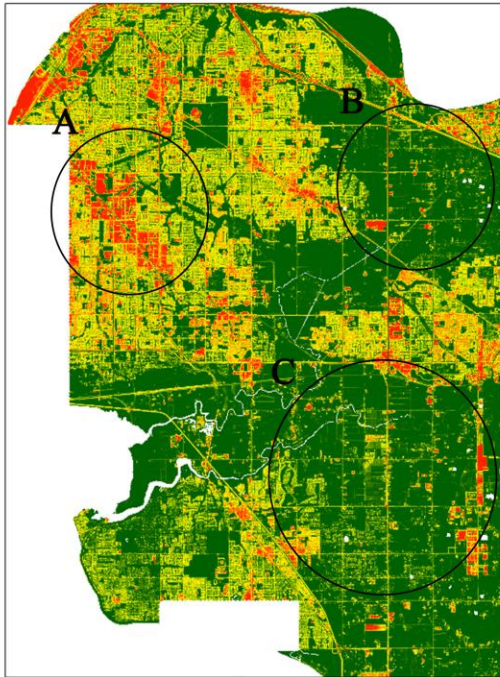
Reference
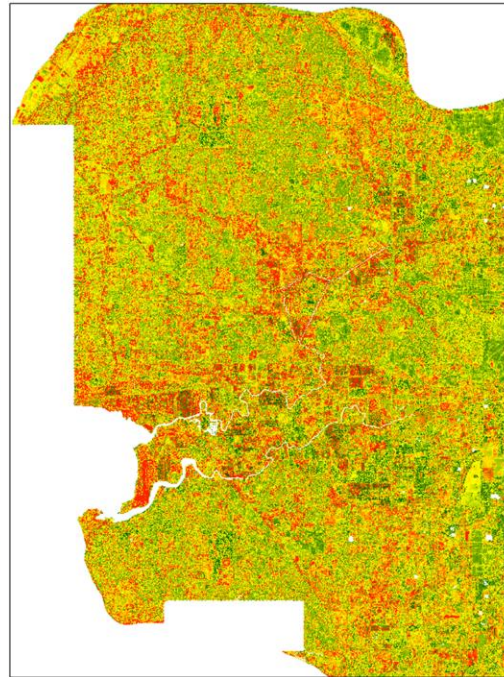
N-Findr

PPI
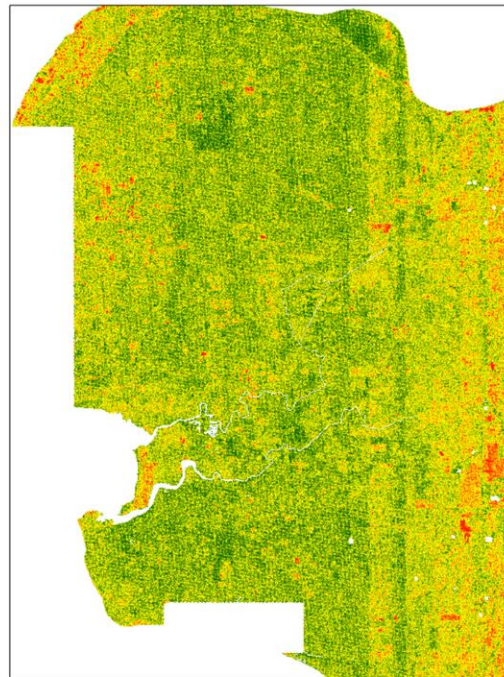
ICA
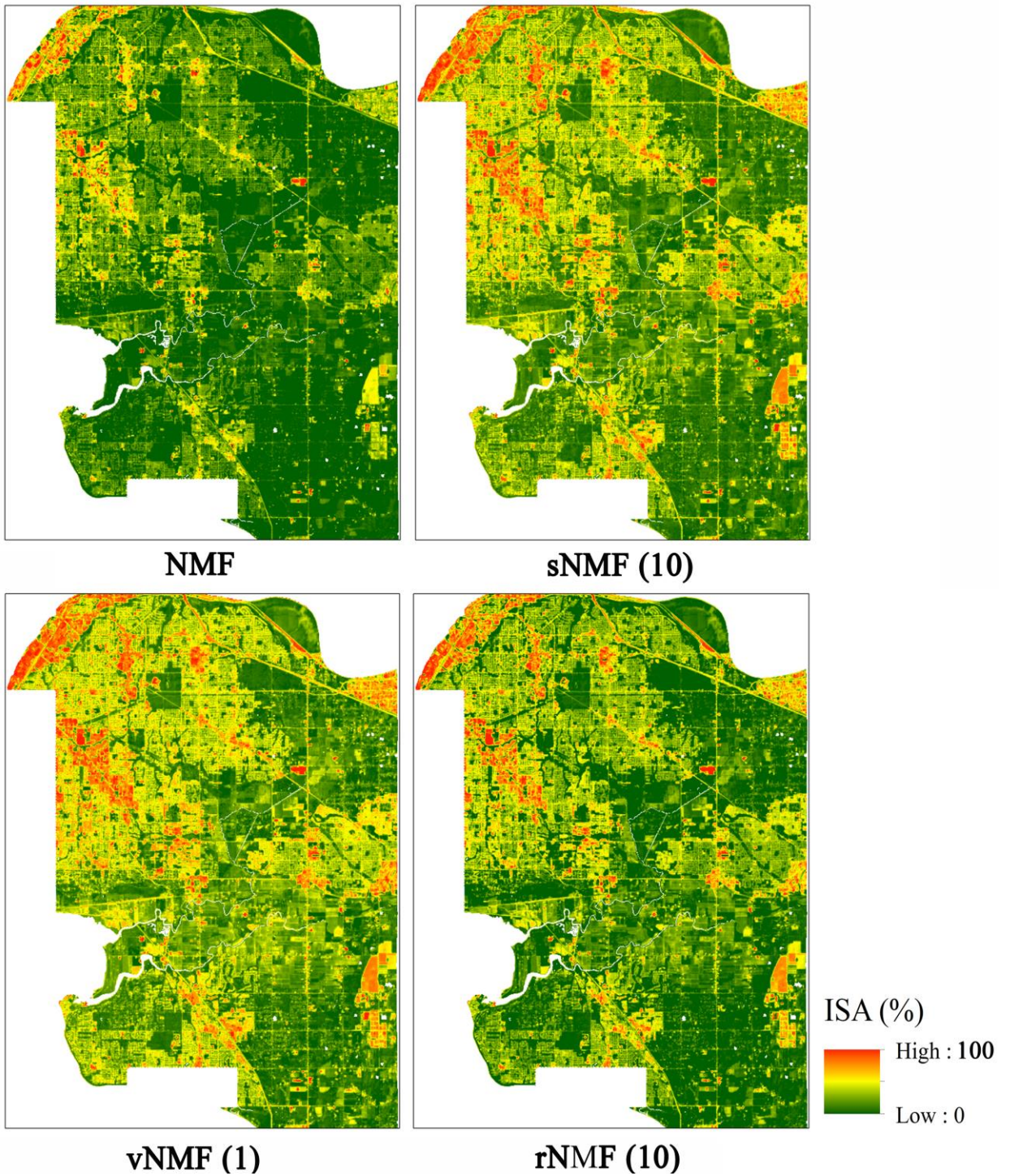
**Figure 2-11: The best impervious surface area abundance maps from each spectral mixture analysis method (EnMAP).**
**(the value in the parenthesis is the used constraint parameter)**

Next we have the Hydice urban image ISA abundance results in Figure 2-12. In the first row of Figure 2-12, the reference ISA abundance map is shown, followed by N-Findr, PPI, and ICA. With a higher spatial resolution than the simulated EnMAP image, the N-Findr and PPI methods were able to produce better results for the Hydice urban image. The biggest inaccuracy in the N-Findr and PPI results is that they tend to underestimate the ISA abundance: the parking lot at the upper right corner (circle A) of the image and a lot of roofs were assigned with low ISA abundance. In the second row of Figure 2-12, the result of the traditional NMF method obtained high ISA abundance areas in the left side of the image more accurately than the geometry- and statistic-based methods. However, the NMF method underestimated the ISA abundance in the bright roof at the upper right corner (circle A) and overestimated the ISA abundance in the vegetation area in the lower part of the image (circle B). The sNMF method also has the two problems found in the NMF, but is less severe than the NMF. The vNMF method predicted the low ISA abundance in circle B, but failed to predict the high ISA abundance in the parking lot in circle A. The rNMF method encountered problems in recognizing the low ISA abundance in circle B and high ISA abundance in circle A.
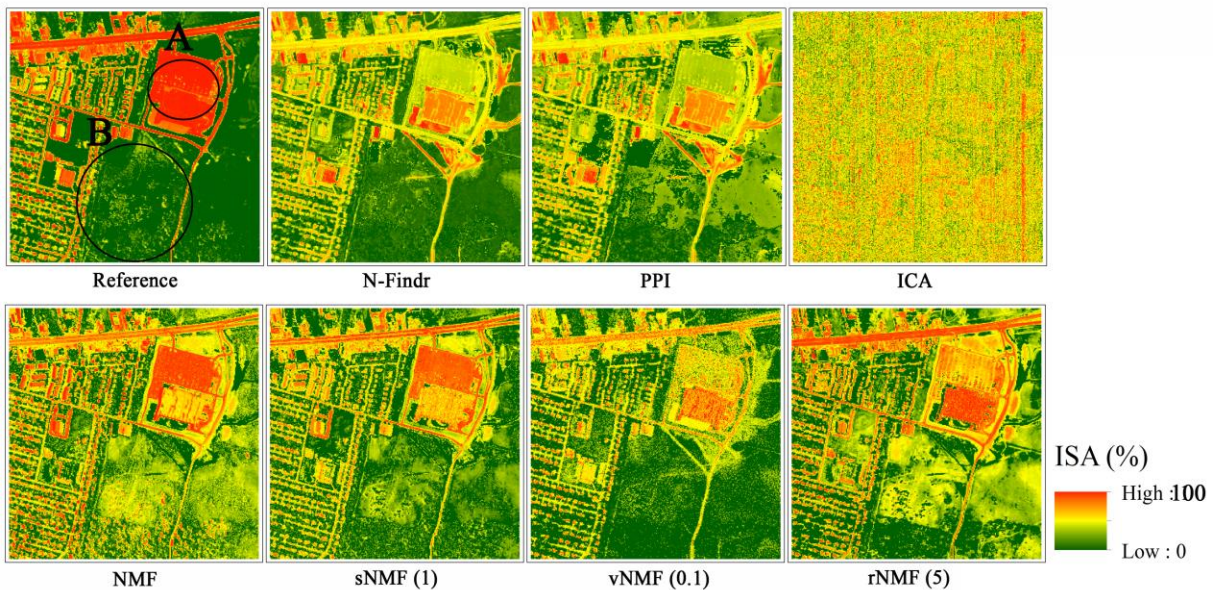


**Figure 2-12: The best impervious surface area abundance maps from each spectral mixture analysis method (Hydice).**
**(the value in the parenthesis is the used constraint parameter)**

An interesting difference found in the ISA abundance results between the two study images is that the Hydice urban image is more sensitive to the changes of different constraint parameters than the simulated EnMAP image. In the experiments of Hydice urban image using the same constrained method, the different constraint parameter values can cause more than 0.4 difference in the correlation coefficient of the linear regression models between the predicted and reference ISA abundances. This difference is less than 0.1 in the experiments of the simulated EnMAP image.

## 2.5.2    ISA classification maps

In the end, the classification results are shown in Figure 2-13 for the simulated EnMAP image and in Figure 2-14 for the Hydice urban image. Begin with the results from the simulated EnMAP image. Only three constrained NMF ISA classification maps were shown for each image, corresponding to the sNMF, vNMF, and rNMF results with the best classification overall accuracy. In the first two rows of Figure 2-13, the reference ISA classification map is shown. Naturally, based on the poor ISA abundance maps from the N-Findr, PPI, and ICA methods, their classification results were also poor. The result from the traditional NMF method appears to be reasonable, except for two inaccuracies: in the Middle Eastern part of the map, the dense ISA was classified as pervious (circle A); and in the Southeastern part of the map a large agricultural bare soil was classified as impervious (circle B). The sNMF, vNMF, and rNMF methods also have the two problems and they are very similar with the NMF results, with only minor differences.
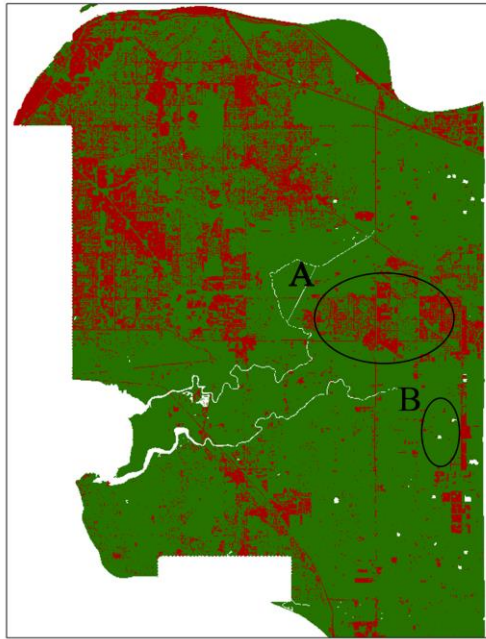
Reference
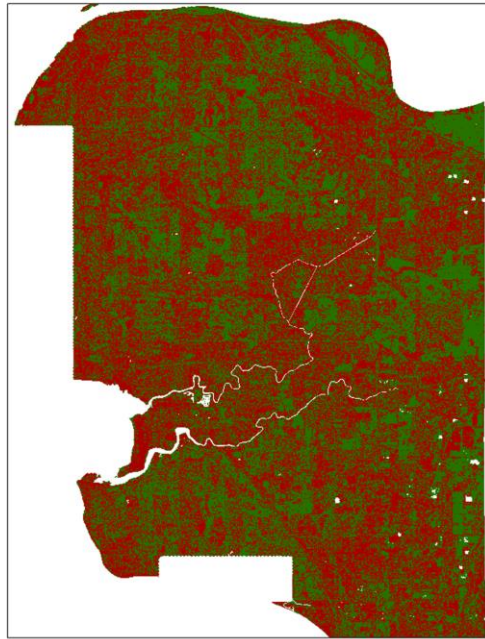
N-Findr
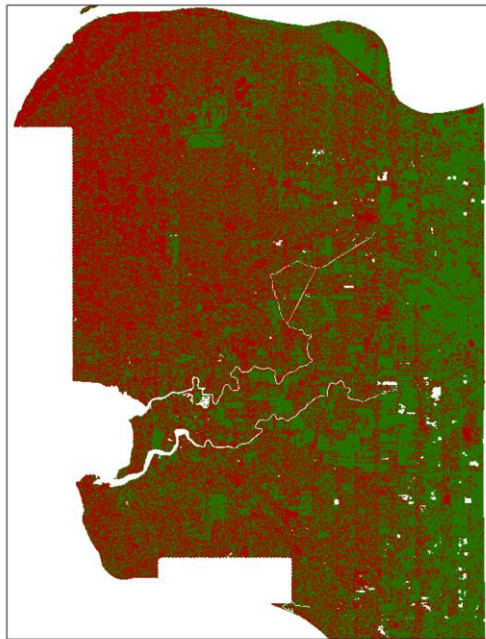
PPI
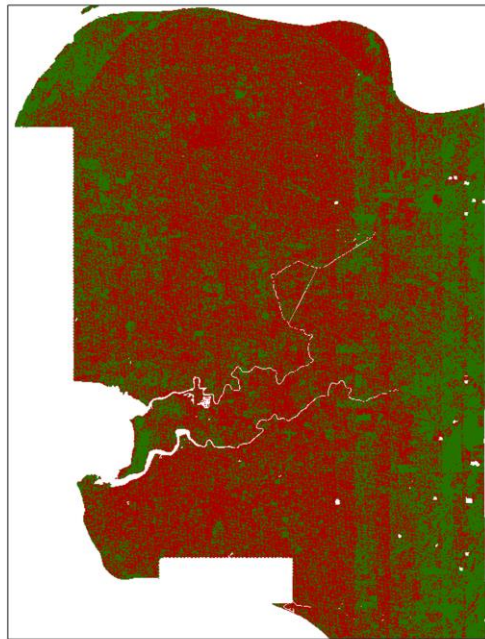
ICA
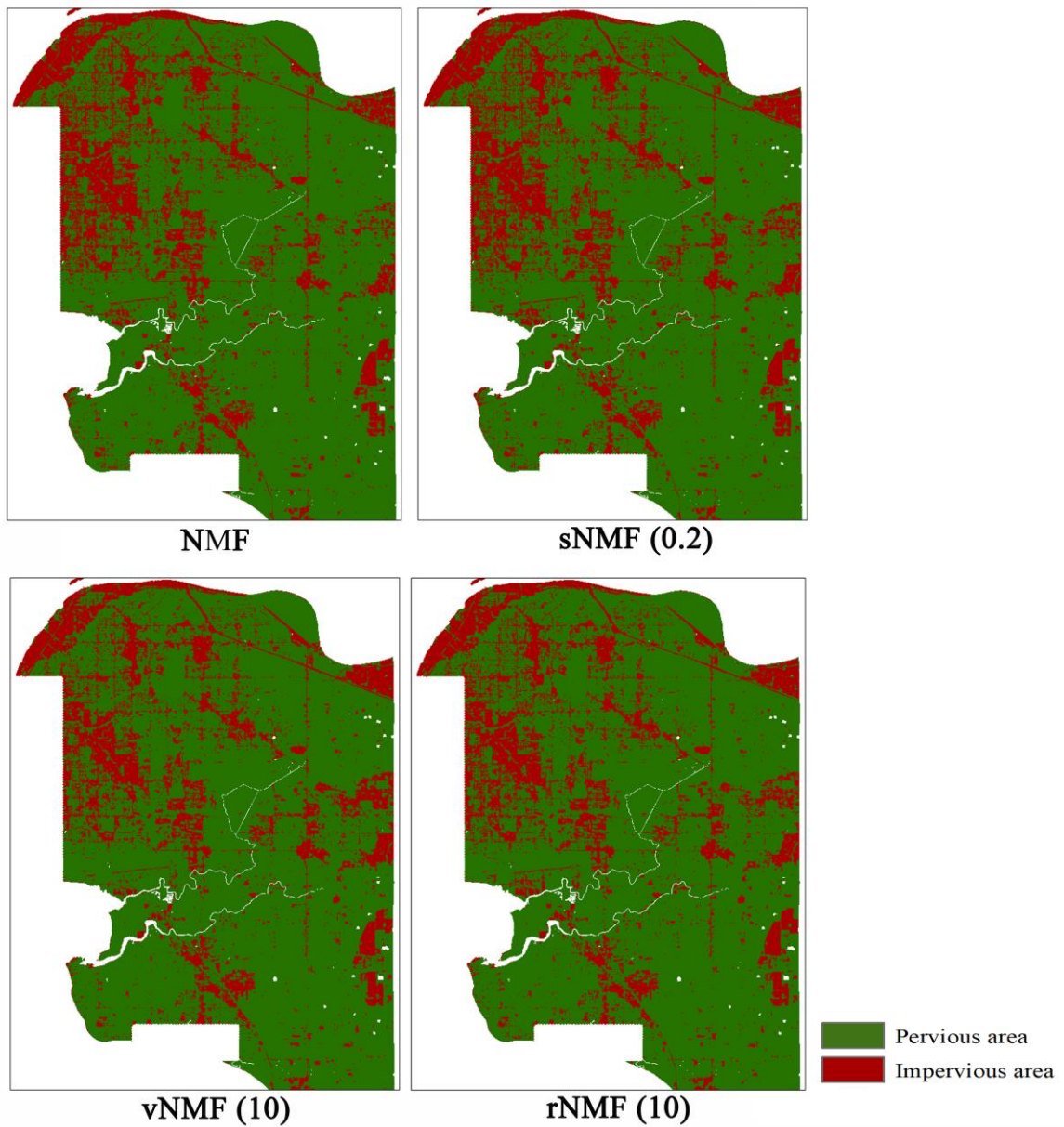
**Figure 2-13: The best impervious surface area classification maps from each spectral mixture analysis method (EnMAP).**
**(the value in the parenthesis is the used constraint parameter)**

In the case of the Hydice urban image, Figure 2-14 shows the reference ISA classification map at the left, followed by N-Findr, PPI, and ICA. The N-Findr and PPI results misclassified the parking lot in the upper right part (circle A) of the image as pervious. The traditional NMF method correctly classified the parking lot. However, it misclassified some part of the vegetation in the lower right part (circle B) of the image as

impervious. The three constrained NMF methods showed better classification results than the traditional NMF method in circle A and B.
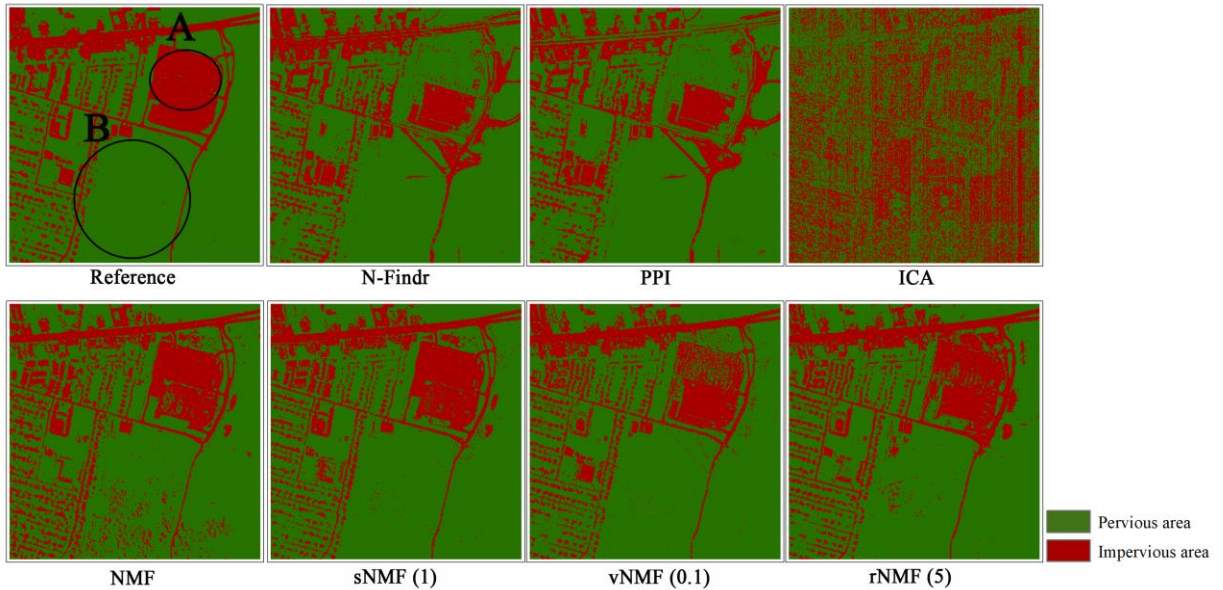


**Figure 2-14: The best impervious surface area abundance maps from each spectral mixture analysis method (Hydice).**
**(the value in the parenthesis is the used constraint parameter)**

As mentioned in the above discussion about ISA abundance maps, the results of the Hydice urban image are more sensitive to the constraint parameter values than the results of the simulated EnMAP image. This fact continues in the ISA classification results. In the experiments of the Hydice urban image using the same constrained method, the different constraint parameter values can cause more than a 10% difference in the overall accuracy values (Table 2-4). This difference is usually smaller than 5% in the experiments of the simulated EnMAP image.

## 2.5.3 Computation time

Lastly, the processing times in seconds for the NMF-based hyperspectral SMA methods are shown in Table 2-5. There are separately 350555 and 94249 pixels in the simulated EnMAP image and the Hydice urban image that need to be processed. The traditional NMF method takes more time than most of the constraint NMF methods for both images. With more definitions on the objective matrix, the constrained NMF methods shrink

down the solution space and locate the final results sooner. The constraint NMF methods (e.g. sNMF (0.7) and some rNMF experiments) that need significantly more processing time usually have a slow convergence and this may affect their results in a disadvantageous way.

**Table 2-7: Processing times in seconds.**

| Method | Time (EnMAP) | Time (Hydice) |
|---|---|---|
| NMF | 117 | 44 |
| sNMF (Equation (2-11)) | 13 | 5 |
| sNMF (0.01) | 50 | 5 |
| sNMF (0.1) | 7 | 9 |
| sNMF (0.2) | 24 | 3 |
| sNMF (0.5) | 5 | 4 |
| sNMF (0.7) | 102 | 107 |
| sNMF (1.0) | 24 | 3 |
| sNMF (5.0) | 7 | 3 |
| sNMF (10.0) | 4 | 5 |
| vNMF (0.01) | 13 | 4 |
| vNMF (0.1) | 58 | 3 |
| vNMF (0.2) | 68 | 5 |
| vNMF (0.5) | 21 | 2 |
| vNMF (0.7) | 3 | 4 |
| vNMF (1.0) | 4 | 4 |
| vNMF (5.0) | 14 | 4 |
| vNMF (10.0) | 5 | 3 |
| rNMF (Equation (2-15)) | 222 | - |
| rNMF (0.01) | 280 | 80 |
| rNMF (0.1) | 40 | 5 |
| rNMF (0.2) | 98 | 7 |
| rNMF (0.5) | 101 | 3 |
| rNMF (0.7) | 4 | 4 |
| rNMF (1.0) | 32 | 3 |
| rNMF (5.0) | 118 | 3 |
| rNMF (10.0) | 60 | 3 |

## 2.6 Conclusion

The future EnMAP satellite will complement current airborne and spaceborne hyperspectral sensors by providing timely and global hyperspectral data. In order to assess the designed EnMAP sensor and offer possible application and processing methods for the future use of EnMAP images, Chapter 2 conducted series experiments on spectral unmixing for ISA mapping. Two geometry-based methods (N-Findr, PPI), one statistic-based method (ICA), and four algebra-based methods were applied to generate ISA abundance and classification maps from the simulated EnMAP image in Surrey, BC, Canada. In addition, the widely used Hydice urban image was processed using the same methods as comparison. We analyzed the ISA mapping results from the aspects of different images and different endmember extraction/hyperspectral SMA methods. The four research questions in 2.2 are correspondingly addressed below.

(1) The EnMAP imagery has a medium spatial resolution of 30m and the pure pixel is limited. The traditional geometry- and statistic-based endmember extraction methods have difficulty locating the non-existing endmembers from the image. In this way, most of the current EnMAP studies have used the endmembers from existing references or other high spatial resolution images. Although the carefully selected endmembers from other sources worked well with the EnMAP image, it should be noted that the existing endmembers are not always available. Using the endmembers derived from the EnMAP image itself is still the most reproducible method. We found that the algebra-based NMF hyperspectral SMA methods can successfully work with the simulated EnMAP image for ISA mapping. The ISA abundance maps derived from the simulated EnMAP image were compared with the reference abundance map using correlation coefficients and MAD values. The resulting correlation coefficients reached 0.7 and the MAD values reached 0.14. These results are as good as the up-to-date ISA studies (Fan, Fan and Weng 2015, Guo, Lu and Kuang 2017), which often obtained help from the additional spectral indices and multi-date or multi-sensor data. From the abundance maps, the pervious/impervious classification maps were generated using thresholds derived from the linear regression models, which achieves more than 85% accuracy when compared with the reference map.

(2) The linear regression results from the airborne high spatial resolution hyperspectral Hydice urban image are better than the results from the simulated EnMAP image. The ISA classification results for the Hydice urban image also provided better results than the simulated EnMAP image. However, the advantages of the Hydice image are limited. Further, considering the different image sizes and land cover heterogeneities between the two images, the potential in the simulated EnMAP image should not be underestimated.

(3) The abundance maps directly obtained from the algebra-based methods underestimated the ISA abundance greatly. It is necessary to perform a linear normalization on the obtained abundance maps to achieve reliable results.

 (4) In the case of the simulated EnMAP image, among the four experimented NMF-based methods (NMF, sNMF, vNMF, and rNMF), the constrained methods did not perform better than the traditional method. Further, the ISA mapping results are not very sensitive to the constraint parameter values for the three constrained NMF methods. When it comes to the Hydice urban image, the NMF-based methods still outperforms the geometry-based and statistical methods despite the fine spatial resolution. The constrained NMF methods provides better ISA results than the traditional NMF method. The Hydice urban image is very sensitive to the constraint parameter values for the three experimented constrained NMF methods. Since the decisions on the constraint parameter values are still not fully understood, images that are sensitive to these values may need extra attempts to obtain the reliable results.

(5) This study shows the potential of using the medium spatial resolution spaceborne EnMAP image for large area ISA mapping in the future. This new hyperspectral instrument provides new economic perspectives toward ISA monitoring, in addition to the currently available multispectral Landsat and more expensive airborne data. Although with the proposed processing procedure, the ISA mapping results from the simulated EnMAP image had similar accuracy as the up-to-date ISA mapping literatures, the method in this study does not need additional data like the current literatures.

# References

Arngren, M., M. N. Schmidt & J. Larsen (2011) Unmixing of hyperspectral images using Bayesian non-negative matrix factorization with volume prior. *Journal of Signal Processing Systems,* 65**,** 479-496.

Arnold Jr, C. L. & C. J. Gibbons (1996) Impervious surface coverage: the emergence of a key environmental indicator. *Journal of the American planning Association,* 62**,** 243-258.

Barnsley, M. J., J. J. Settle, M. A. Cutter, D. R. Lobb & F. Teston (2004) The PROBA/CHRIS mission: A low-cost smallsat for hyperspectral multiangle observations of the earth surface and atmosphere. *IEEE Transactions on Geoscience and Remote Sensing,* 42**,** 1512-1520.

Bioucas-Dias, J. M., A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader & J. Chanussot (2012) Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE journal of selected topics in applied earth observations and remote sensing,* 5**,** 354-379.

Boardman, J. W. 1993. Automating spectral unmixing of AVIRIS data using convex geometry concepts. In *Summaries 4th Annu. JPL Airborne Geoscience Workshop*, 11-14. JPL Publication 93–26.

Boardman, J. W., F. A. Kruse & R. O. Green (1995) Mapping target signatures via partial unmixing of AVIRIS data.

Braun, A. C., U. Weidner & S. Hinz (2012) Classification in high-dimensional feature spaces—Assessment using SVM, IVM and RVM with focus on simulated EnMAP data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 5**,** 436-443.

Chang, C.-I. & Q. Du (2004) Estimation of number of spectrally distinct signal sources in hyperspectral imagery. *IEEE Transactions on geoscience and remote sensing,* 42**,** 608-619.

Chang, C.-I., C.-C. Wu & H.-M. Chen (2010) Random pixel purity index. *Geoscience and Remote Sensing Letters, IEEE,* 7**,** 324-328.

Cichocki, A., R. Zdunek, A. H. Phan & S.-i. Amari. 2009. *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons.

Clasen, A., B. Somers, K. Pipkins, L. Tits, K. Segl, M. Brell, B. Kleinschmit, D. Spengler, A. Lausch & M. Förster (2015) Spectral unmixing of forest crown components at

close range, airborne and simulated Sentinel-2 and EnMAP spectral imaging scale. *Remote Sensing,* 7**,** 15361-15387.

Cocks, T., R. Jenssen, A. Stewart, I. Wilson & T. Shields. 1998. The HyMapTM airborne hyperspectral sensor: the system, calibration and performance. In *Proceedings of the 1st EARSeL workshop on Imaging Spectroscopy*, 37-42. EARSeL.

Computational Intelligence Group, U. o. t. B. C. U. E. Endmember Induction Algorithms (EIAs) toolbox. [http://www.ehu.es/computationalintelligence/index.php/Endmember_Induction_Algorithms](http://www.ehu.es/computationalintelligence/index.php/Endmember_Induction_Algorithms).

Corson, M. R., D. R. Korwan, R. L. Lucke, W. A. Snyder & C. O. Davis. 2008. The hyperspectral imager for the coastal ocean (HICO) on the international space station. In *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*, IV-101-IV-104. IEEE.

Craig, M. D. (1994) Minimum-volume transforms for remotely sensed data. *IEEE Transactions on Geoscience and Remote Sensing,* 32**,** 542-552.

Dennison, P. E., K. Q. Halligan & D. A. Roberts (2004) A comparison of error metrics and constraints for multiple endmember spectral mixture analysis and spectral angle mapper. *Remote Sensing of Environment,* 93**,** 359-367.

Dotzler, S., J. Hill, H. Buddenbaum & J. Stoffels (2015) The potential of EnMAP and Sentinel-2 data for detecting drought stress phenomena in deciduous forest communities. *Remote Sensing,* 7**,** 14227-14258.

Dozier, J. (1981) A method for satellite identification of surface temperature fields of subpixel resolution. *Remote Sensing of environment,* 11**,** 221-229.

Elith, J., J. R. Leathwick & T. Hastie (2008) A working guide to boosted regression trees. *Journal of Animal Ecology,* 77**,** 802-813.

Fan, F., W. Fan & Q. Weng (2015) Improving urban impervious surface mapping by linear spectral mixture analysis and using spectral indices. *Canadian Journal of Remote Sensing,* 41**,** 577-586.

Fassnacht, F. E., H. Weinacker & B. Koch. 2011. Automatic forest area extraction from imaging spectroscopy data using an extended NDVI. In *7th EARSeL Workshop on Imaging Spectroscopy, Edinburgh, Scotland*.

Févotte, C. & N. Dobigeon (2015) Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization. *IEEE Transactions on Image Processing,* 24**,** 4810-4819.

Gower, J. F. R., G. A. Borstad, C. D. Anger & H. R. Edel (1992) CCD-based imaging spectroscopy for remote sensing: the FLI and CASI programs. *Canadian journal of remote sensing,* 18**,** 199-208.

Guanter, L., H. Kaufmann, K. Segl, S. Foerster, C. Rogass, S. Chabrillat, T. Kuester, A. Hollstein, G. Rossner & C. Chlebek (2015) The EnMAP spaceborne imaging spectroscopy mission for earth observation. *Remote Sensing,* 7**,** 8830-8857.

Guo, W., D. Lu & W. Kuang (2017) Improving fractional impervious surface mapping performance through combination of DMSP-OLS and MODIS NDVI data. *Remote Sensing,* 9**,** 375.

Heldens, W., U. Heiden, T. Esch, E. Stein & A. Müller (2011) Can the future EnMAP mission contribute to urban applications? A literature survey. *Remote Sensing,* 3**,** 1817-1846.

Hoyer, P. O. (2004) Non-negative matrix factorization with sparseness constraints. *Journal of machine learning research,* 5**,** 1457-1469.

Hyvärinen, A. & E. Oja (2000) Independent component analysis: algorithms and applications. *Neural networks,* 13**,** 411-430.

Karatzoglou, A., D. Meyer & K. Hornik (2005) Support vector machines in R.

Kaufmann, H., K. Segl, S. Chabrillat, S. Hofer, T. Stuffler, A. Mueller, R. Richter, G. Schreier, R. Haydn & H. Bach. 2006. EnMAP a hyperspectral sensor for environmental mapping and analysis. In *Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006. IEEE International Conference on*, 1617-1619. IEEE.

Lee, A., F. Caron, A. Doucet & C. Holmes (2010) A hierarchical Bayesian framework for constructing sparsity-inducing priors. *arXiv preprint arXiv:1009.1914*.

Lehnert, L. W., H. Meyer, N. Meyer, C. Reudenbach & J. Bendix (2014) A hyperspectral indicator system for rangeland degradation on the Tibetan Plateau: A case study towards spaceborne monitoring. *Ecological Indicators,* 39**,** 54-64.

Leitão, P. J., M. Schwieder, S. Suess, A. Okujeni, L. S. Galvão, S. v. d. Linden & P. Hostert (2015) Monitoring natural ecosystem and ecological gradients: Perspectives with EnMAP. *Remote Sensing,* 7**,** 13098-13119.

Liaw, A. & M. Wiener (2002) Classification and regression by randomForest. *R news,* 2**,** 18-22.

Liu, W., K. C. Seto, E. Y. Wu, S. Gopal & C. E. Woodcock (2004) ART-MMAP: A neural network approach to subpixel classification. *Geoscience and Remote Sensing, IEEE Transactions on,* 42**,** 1976-1983.

Locherer, M., T. Hank, M. Danner & W. Mauser (2015) Retrieval of seasonal leaf area index from simulated EnMAP data through optimized LUT-based inversion of the PROSAIL model. *Remote Sensing,* 7**,** 10321-10346.

Lu, X., H. Wu, Y. Yuan, P. Yan & X. Li (2013) Manifold regularized sparse NMF for hyperspectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing,* 51**,** 2815-2826.

Luo, B., J. Chanussot, S. Douté & L. Zhang (2013) Empirical automatic estimation of the number of endmembers in hyperspectral images. *IEEE Geoscience and Remote Sensing Letters,* 10**,** 24-28.

Ma, W.-K., J. M. Bioucas-Dias, T.-H. Chan, N. Gillis, P. Gader, A. J. Plaza, A. Ambikapathi & C.-Y. Chi (2014) A signal processing perspective on hyperspectral unmixing: Insights from remote sensing. *IEEE Signal Processing Magazine,* 31**,** 67-81.

Malec, S., D. Rogge, U. Heiden, A. Sanchez-Azofeifa, M. Bachmann & M. Wegmann (2015) Capability of spaceborne hyperspectral EnMAP mission for mapping fractional cover for soil erosion modeling. *Remote Sensing,* 7**,** 11776-11800.

Marcinkowska-Ochtyra, A., B. Zagajewski, A. Ochtyra, A. Jarocińska, B. Wojtuń, C. Rogass, C. Mielke & S. Lavender (2017) Subalpine and alpine vegetation classification based on hyperspectral APEX and simulated EnMAP images. *International Journal of Remote Sensing,* 38**,** 1839-1864.

McKinney, M. L. (2008) Effects of urbanization on species richness: a review of plants and animals. *Urban ecosystems,* 11**,** 161-176.

Miao, L. & H. Qi (2007) Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing,* 45**,** 765-777.

Mielke, C., N. K. Boesche, C. Rogass, H. Kaufmann, C. Gauert & M. de Wit (2014) Spaceborne Mine Waste Mineralogy Monitoring in South Africa, Applications for Modern Push-Broom Missions: Hyperion/OLI and EnMAP/Sentinel-2. *Remote Sensing,* 6**,** 6790-6816.

Nascimento, J. M. & J. M. Dias (2005) Does independent component analysis play a role in unmixing hyperspectral data? *IEEE Transactions on Geoscience and Remote Sensing,* 43**,** 175-187.

Okujeni, A., S. van der Linden & P. Hostert (2015) Extending the vegetation–impervious–soil model using simulated EnMAP data and machine learning. *Remote Sensing of Environment,* 158**,** 69-80.

Okujeni, A., S. van der Linden, L. Tits, B. Somers & P. Hostert (2013) Support vector regression and synthetically mixed training data for quantifying urban land cover. *Remote Sensing of Environment,* 137**,** 184-197.

Oldenburg, C., S. Schmidtlein & H. Feilhauer. AutoPLSR: Manual for Application: AutoPLSR (1. x), EnMAP-Box Documentation.

Paatero, P. & U. Tapper (1994) Positive matrix factorization: A non‑ negative factor model with optimal utilization of error estimates of data values. *Environmetrics,* 5**,** 111-126.

Pearlman, J. S., P. S. Barry, C. C. Segal, J. Shepanski, D. Beiso & S. L. Carman (2003) Hyperion, a space-based imaging spectrometer. *IEEE Transactions on Geoscience and Remote Sensing,* 41**,** 1160-1173.

Plaza, A., P. Martínez, R. Pérez & J. Plaza (2004) A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data. *Geoscience and Remote Sensing, IEEE Transactions on,* 42**,** 650-663.

Powell, R. L., D. A. Roberts, P. E. Dennison & L. L. Hess (2007) Sub-pixel mapping of urban land cover using multiple endmember spectral mixture analysis: Manaus, Brazil. *Remote Sensing of Environment,* 106**,** 253-267.

Qian, Y., S. Jia, J. Zhou & A. Robles-Kelly (2011) Hyperspectral unmixing via $ L\_ \{1/2\} $ sparsity-constrained nonnegative matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing,* 49**,** 4282-4297.

Rickard, L. J., R. W. Basedow, E. F. Zalewski, P. R. Silverglate & M. Landers. 1993. HYDICE: An airborne system for hyperspectral imaging. In *Optical Engineering and Photonics in Aerospace Sensing*, 173-179. International Society for Optics and Photonics.

Ridd, M. K. (1995) Exploring a VIS (vegetation-impervious surface-soil) model for urban ecosystem analysis through remote sensing: comparative anatomy for cities†. *International journal of remote sensing,* 16**,** 2165-2185.

Roberts, D. A., P. E. Dennison, M. E. Gardner, Y. Hetzel, S. L. Ustin & C. T. Lee (2003) Evaluation of the potential of Hyperion for fire danger assessment by comparison to the Airborne Visible/Infrared Imaging Spectrometer. *Geoscience and Remote Sensing, IEEE Transactions on,* 41**,** 1297-1310.

Roberts, D. A., M. Gardner, R. Church, S. Ustin, G. Scheer & R. Green (1998a) Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. *Remote Sensing of Environment,* 65**,** 267-279.

Roberts, D. A., M. Gardner, R. Church, S. Ustin, G. Scheer & R. O. Green (1998b) Mapping chaparral in the Santa Monica Mountains using multiple endmember spectral mixture models. *Remote Sensing of Environment,* 65**,** 267-279.

Rogge, D., M. Bachmann, B. Rivard & J. Feng (2012) Spatial sub-sampling using local endmembers for adapting OSP and SSEE for large-scale hyperspectral surveys. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 5**,** 183-195.

Rogge, D., B. Rivard, K. Segl, B. Grant & J. Feng (2014) Mapping of NiCu–PGE ore hosting ultramafic rocks using airborne and simulated EnMAP hyperspectral imagery, Nunavik, Canada. *Remote sensing of environment,* 152**,** 302-317.

Rogge, D. M., B. Rivard, J. Zhang & J. Feng (2006) Iterative spectral unmixing for optimizing per-pixel endmember sets. *IEEE Transactions on Geoscience and Remote Sensing,* 44**,** 3725-3736.

Schachtner, R., G. Pöppel, A. M. Tomé & E. W. Lang. 2009. Minimum Determinant Constraint for Non-negative Matrix Factorization. In *Ica*, 106-113. Springer.

Schwieder, M., P. J. Leitão, S. Suess, C. Senf & P. Hostert (2014) Estimating fractional shrub cover using simulated EnMAP data: A comparison of three machine learning regression techniques. *Remote Sensing,* 6**,** 3427-3445.

Segl, K., L. Guanter, C. Rogass, T. Kuester, S. Roessner, H. Kaufmann, B. Sang, V. Mogulsky & S. Hofer (2012) EeteS—The EnMAP end-to-end simulation tool. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 5**,** 522-530.

Siegmann, B., T. Jarmer, F. Beyer & M. Ehlers (2015) The potential of pan-sharpened EnMAP data for the assessment of wheat LAI. *Remote Sensing,* 7**,** 12737-12762.

Steinberg, A., S. Chabrillat, A. Stevens, K. Segl & S. Foerster (2016) Prediction of common surface soil properties based on Vis-NIR airborne and simulated EnMAP imaging spectroscopy data: Prediction accuracy and influence of spatial resolution. *Remote Sensing,* 8**,** 613.

Suess, S., S. van der Linden, A. Okujeni, P. J. Leitão, M. Schwieder & P. Hostert (2015) Using class probabilities to map gradual transitions in shrub vegetation from simulated EnMAP data. *Remote Sensing,* 7**,** 10668-10688.

Tipping, M. E. & A. C. Faul. 2003. Fast marginal likelihood maximisation for sparse Bayesian models. In *Aistats*.

Vane, G., R. O. Green, T. G. Chrien, H. T. Enmark, E. G. Hansen & W. M. Porter (1993) The airborne visible/infrared imaging spectrometer (AVIRIS). *Remote sensing of environment,* 44**,** 127-143.

Voogt, J. A. & T. R. Oke (2003) Thermal remote sensing of urban climates. *Remote sensing of environment,* 86**,** 370-384.

Wang, J. & C.-I. Chang (2006) Applications of independent component analysis in endmember extraction and abundance quantification for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing,* 44**,** 2601-2616.

Wehrens, R. & B. H. Mevik (2007) The pls package: principal component and partial least squares regression in R.

Weng, Q. (2012) Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends. *Remote Sensing of Environment,* 117**,** 34-49.

Winter, M. E. 1999. N-FINDR: an algorithm for fast autonomous spectral end-member determination in hyperspectral data. In *SPIE's International Symposium on Optical Science, Engineering, and Instrumentation*, 266-275. International Society for Optics and Photonics.

Wolch, J. R., J. Byrne & J. P. Newell (2014) Urban green space, public health, and environmental justice: The challenge of making cities 'just green enough'. *Landscape and Urban Planning,* 125**,** 234-244.

Wu, C. & A. T. Murray (2003) Estimating impervious surface distribution by spectral mixture analysis. *Remote sensing of Environment,* 84**,** 493-505.

Xi, H., M. Hieronymi, R. Röttgers, H. Krasemann & Z. Qiu (2015) Hyperspectral differentiation of phytoplankton taxonomic groups: a comparison between using remote sensing reflectance and absorption spectra. *Remote Sensing,* 7**,** 14781-14805.

Yokoya, N., J. C.-W. Chan & K. Segl (2016) Potential of resolution-enhanced hyperspectral data for mineral mapping using simulated EnMAP and Sentinel-2 images. *Remote Sensing,* 8**,** 172.

Yokoya, N., J. Chanussot & A. Iwasaki (2014) Nonlinear unmixing of hyperspectral data using semi-nonnegative matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing,* 52**,** 1430-1437.

Yokoya, N., T. Yairi & A. Iwasaki (2012) Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion. *IEEE Transactions on Geoscience and Remote Sensing,* 50**,** 528-537.

Zhou, Y., J. Peng & C. P. Chen (2015) Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing,* 53**,** 1082-1095.

Zhu, F., Y. Wang, S. Xiang, B. Fan & C. Pan (2014) Structured sparse method for hyperspectral unmixing. *ISPRS Journal of Photogrammetry and Remote Sensing,* 88**,** 101-118.

Zhu, J. & T. Hastie (2005) Kernel logistic regression and the import vector machine. *Journal of Computational and Graphical Statistics,* 14**,** 185-205.

# 3 Constrained nonnegative tensor factorization for spectral mixture analysis of hyperspectral image

The nonnegative matrix factorization (NMF) discussed in Chapter 2 adopts a linear algebra solution that decomposes the original data into one endmember and one abundance matrices. Although the linear algebra solution in the NMF-based spectral mixture analysis (SMA) methods loosens the requirement of the existing pure pixels in the original image, this method only considers the characteristics of one individual pixel, and the spatial information in the small neighborhood of the pixel is discarded. In detail, since NMF is based on linear algebra, a necessary preprocess of the NMF-based hyperspectral SMA methods is to convert the 3D hyperspectral data cubes into 2D matrices, where rows represent observations and columns represent spectral bands. By this conversion, the spatial information in the relative positons of the pixels is lost. This limitation can be found in many classical hyperspectral processes (e.g. PCA (principal component analysis)). With the emergence of multilinear algebra, the tensorial representation of hyperspectral imagery has become popular. A hyperspectral image starts to be treated as a third-order tensor with two spatial dimensions and one spectral dimension. Recently, the tensor-based hyperspectral methods have gradually gained attention in remote sensing community. Some attempts involving feature extraction (Ren et al. 2017, Liu et al. 2017), data compression (Fang, He and Lin 2017, Du et al. 2017), dimension reduction (DR) (Gao et al. 2015, Deng et al. 2018), and SMA (Veganzones et al. 2016, Qian et al. 2017) were made. As a new direction in remote sensing, the tensor-based SMA methods need to be further studied.

## 3.1 Background

The idea of tensor-based hyperspectral SMA has been brought up as early as 2007 (Zhang et al. 2007, Zhang et al. 2008). However, the implementation of this idea was impeded by the lack of physical interpretation in the traditional tensor factorization methods. Basically, the linear SMA must have two characteristics: (1) the number of

endmembers should be fixed during the calculation, and (2) the original image should be represented as a sum of component tensors. The CANDECOMP/PARAFAC (CP) decomposition decomposes a given tensor into the sum of the minimum rank-one tensors (Figure 1-8). The minimum number of the component rank-one tensors is called the tensor rank. The determination of the tensor rank is not straightforward (Kolda and Bader 2009), and does not reflect the number of endmembers. Therefore, it is difficult to use the CP decomposition to deduce the endmember and corresponding abundance matrix. On the other hand, the Tucker decomposition defines a given tensor as a core tensor multiplied by one matrix along each mode (Figure 1-8). The decomposing manner fails to divide the tensor into a sum of components tensor, weakening its link to linear SMA. In addition, similar to the CP decomposition, no straightforward method is available to determine the size of the core tensor using the Tucker decomposition. Although the spectral mode rank in the core tensor can be set to the number of endmembers, the derived endmembers will be orthogonal to each other, which conflicts with the nature of endmembers. Thus, the CP decomposition and Tucker decomposition cannot be directly used as linear SMA (Equation 3-1).

It can be observed that the CP decomposition and the Tucker decomposition each possesses one critical characteristic in the linear SMA analysis. A hybrid tensor decomposition that combines the CP and Tucker decompositions is a promising direction to conform the form of SMA. Research has been done in this direction (Kolda and Bader 2009). One popular hybrid tensor decomposition method is block term decomposition (BTD) (De Lathauwer 2008). The BTD decomposes the given tensor as a sum of component tensors, where each component tensor is the $k$-mode product of the core tensor and factor matrices (Figure 3-1). The BTD has flexible rules for the component tensors, as the component tensors are not restricted to be rank-one tensors like they are in the CP decomposition. Then, the number of the component tensors can be easily set as the number of endmembers. At the same time, the BTD adopts a summation form of the component tensors. For a third-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$, its BTD expression is as below:

$$\mathcal{X} = \sum_{r=1}^{3} \mathcal{G}_r \times_1 A_r \times_2 B_r \times_3 C_r , \qquad\qquad (3\text{-}1)$$
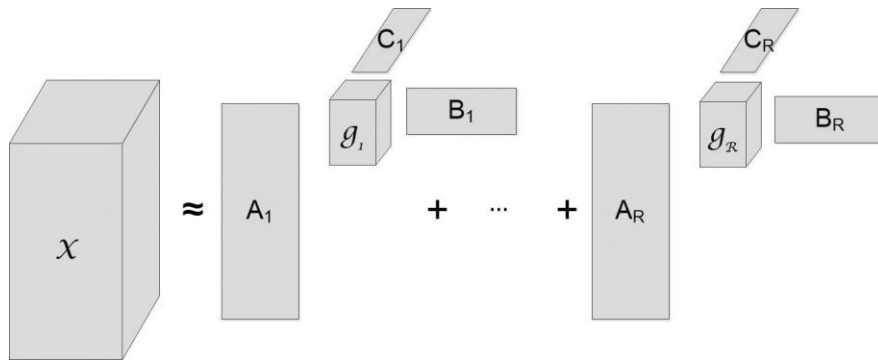
**Figure 3-1: BTD (Qian et al. 2017).**

Inspired by the BTD, Qian, et al., first realized tensor-based SMA through a new form of tensor decomposition: matrix-vector nonnegative tensor decomposition (MVNTF) (Qian et al. 2017). By reforming the decomposition manner of the component tensors in BTD, the MVNTF achieves an analogous form to the linear SMA. Specifically, the MVNTF removes the core tensor in the component tensors and factorizes each third-order component tensors into the outer product of one matrix and one vector, among which the matrix represents the abundance, and the vector represents endmember (Figure 3-2).
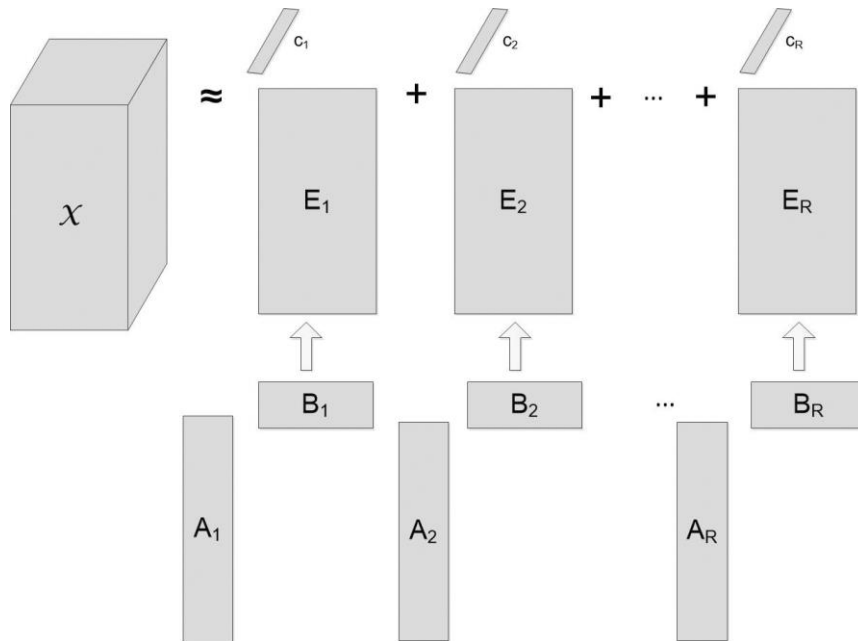


**Figure 3-2: MVNTF (Qian et al. 2017).**

The experiments of MVNTF in SMA confirmed its potential in endmember extraction, but the derived abundance matrix is not fully analyzed (Qian et al. 2017). Another concern about the MVNTF is that it only has one nonnegative constraint, leading to a large solution space. Having too few constraints during decomposition may cause inefficient conversion. Two common disadvantages of the inefficient conversion are finding the unwanted local minimum and increasing the processing time. This problem also occurs in the traditional NMF method and additional constraints were proposed to mitigate the problem. Thus, this research works on integrating similar additional constraints from the NMF-based SMA in the MVNTF-based SMA. Under the construction of MVNTF-based SMA, this research integrated three additional constraints in the cost function, in order to improve the performance of SMA. The three constraints are a sparseness constraint on the abundance map, a minimum volume constraint on endmember, and a nonlinear term on the approximated tensor. These three constraints have been successfully installed into the NMF framework (Bioucas-Dias et al. 2012, Miao and Qi 2007, Arngren, Schmidt and Larsen 2011, Févotte and Dobigeon 2015). This research has found that they also contribute in MVNTF SMA.

## 3.2 Objective

The main objective of Chapter 3 is to upgrade one of the most robust spectral unmixing methods (NMF) to its tensor version and further improve the tensor-based method by incorporating additional constraints that have been previously added to the matrix-based method, to test improvement in the ISA mapping results. Corresponding to the objective, the following research questions are the focus of Chapter 3:

(1) In the two cases of the simulated EnMAP and Hydice urban images, how will the tensor-based SMA methods benefit the ISA mapping, compared to the matrix-based SMA methods?

(2) Can the additional constraints improve the tensor-based SMA methods, as they did for the matrix-based SMA methods?

(3) Are the abundance maps directly obtained from the tensor-based methods usable?

(4) How can the research results help the current ISA mapping study?

## 3.3 Methods

### 3.3.1 Linear spectral mixture analysis

In order to perform SMA with tensor factorization, the intrinsic mechanism of the tensor factorization needs to be analogous to the physical equations of linear spectral mixture:

$$\rho'_\lambda = \sum_{i=1}^{N} f_i * \rho_{i\lambda} + \varepsilon_\lambda \qquad (3\text{-}2)$$

where $\rho_{i\lambda}$ is the $i^{th}$ endmember's reflectivity in band $\lambda$; $f_i$ is the area percentage that $i^{th}$ endmember occupies in the pixel; N is the number of endmembers used in the model; $\varepsilon_\lambda$ is residual. In order to come up with accurate and reliable estimation, two limits are needed: (1) the sum of $f_i$ is one; and (2) $f_i$ is positive for every endmember.

### 3.3.2 Notations and tensor products

In Chapter 1, the basic notations and operations in tensor studies were introduced. Here, I will remind you some of the operations and explain some more notations and one additional tensor operation exclusively used in this Chapter. Matrices are denoted as capital letters, e.g. $A$. Matrices can have subscripts of two multiplied capital letters, e.g. $A_{M \times N}$, representing the sizes of matrices or one capital letter, e.g. $A_R$, representing the locations of matrices if they belong to partitioned matrices [4].

---

[4] In mathematics, a block matrix or a partitioned matrix is a matrix that is interpreted as having been broken into sections called blocks or submatrices.

The outer product is the tensor product of two coordinate vectors. Given the two tensors $\mathcal{A} \in \mathbb{R}^{M_1 \times M_2 \cdots \times M_P}$ and $\mathcal{B} \in \mathbb{R}^{N_1 \times N_2 \cdots \times N_Q}$, the outer product of them is written as $\mathcal{A} \circ \mathcal{B} \in \mathbb{R}^{M_1 \times M_2 \cdots \times M_P \times N_1 \times N_2 \cdots \times N_Q}$. The element is obtained as below:

$$(\mathcal{A} \circ \mathcal{B})_{m_1 m_2 \ldots m_P n_1 n_2 \ldots n_Q} = a_{m_1 m_2 \ldots m_P} b_{n_1 n_2 \ldots n_Q}, \tag{3-3}$$

The Kronecker product of the two matrices $A_{M \times N}$ and $B_{K \times L}$ is $A_{M \times N} \otimes B_{K \times L} \in \mathbb{R}^{MK \times NL}$:

$$A \otimes B = \begin{pmatrix} a_{11}B & \cdots & a_{1N}B \\ \vdots & \ddots & \vdots \\ a_{M1}B & \cdots & a_{MN}B \end{pmatrix}, \tag{3-4}$$

The Khatri-Rao product of the two matrices $A_{M \times N}$ and $B_{K \times N}$ is $A_{M \times N} \odot B_{K \times N} \in \mathbb{R}^{MK \times N}$:

$$A \odot B = [a_1 \otimes b_1 \quad a_2 \otimes b_2 \quad \cdots \quad a_N \otimes b_N], \tag{3-5}$$

where $a_N$ is the $N$th column of matrix $A_{M \times N}$ and $b_N$ is the $N$th column of matrix $B_{M \times N}$. The generalized Khatri-Rao product for the two partitioned matrices $A = [A_1, A_2, \cdots A_R]$ and $B = [B_1, B_2, \cdots B_R]$ results is:

$$A \overline{\odot} B = [A_1 \otimes B_1 \quad A_2 \otimes B_2 \quad \cdots \quad A_R \otimes B_R], \tag{3-6}$$

## 3.3.3    Matrix-vector nonnegative tensor factorization

Earlier studies have stated the potential of using the CP and Tucker decompositions for SMA. However, they have different components from Equation (3-2) and cannot be applied to SMA. The MVNTF method addresses this difficulty by combining the CP and Tucker decompositions, which allows it to work with more complicated data. Figure 3-2 shows the decomposition manner of MVNTF in the case of the third-order tensor. According to Figure 3-2, the corresponding decomposition equation is as below:

$$\mathcal{X} = \sum_{r=1}^{R} A_r \cdot B_r^T \circ c_r = \sum_{r=1}^{R} E_r \circ c_r, \tag{3-7}$$

where $\mathcal{X} \in \mathbb{R}^{M \times N \times K}$ is the original hyperspectral data tensor with $M$ rows, $N$ columns, and $K$ bands; $R$ is the rank of the tensor, defined by the smallest number of component

tensors that generate $X$ as their sum; $A_r \in \mathbb{R}^{M \times L}$ is the decomposed matrix on the row dimension; $B_r \in \mathbb{R}^{N \times L}$ is the decomposed matrix on the column dimension; $L$ is a manually determined column number for $A_r$ and $B_r$; $c_r$ is the decomposed vector on the spectral dimension; $E_r$ is the product of $A_r$ and $B_r$. The resemblance between Equation (3-2) and (3-7) is the foundation of applying MVNTF as SMA. In this way, the $E_r$ represents the abundance matrix and the $c_r$ represents the endmember vector. Despite the nonnegative constraint, another unavoidable constraint for SMA is the sum-to-one constraint on the abundance matrix. It can be attended by a penalty term in the cost function. Thus, the final cost function to solve Equation (3-7) is as below:

$$min_{E,c} \parallel \mathcal{X} - \sum_{r=1}^{R} F_r \circ c_r \parallel_F^2 + \delta \parallel \sum_{r=1}^{R} F_r - J_{M \times N} \parallel_F^2,$$

$$(3\text{-}8)$$

$$s.t. A_r, B_r, c_r \geqslant 0,$$

where $\parallel \parallel_F$ is the third-order tensor Frobenius norm defined as $\parallel \mathcal{X} \parallel_F = \sqrt{\sum_m \sum_n \sum_k x_{mnk}^2}$; the first Frobenius norm is the reconstruction error and the second Frobenius norm reflects the sum-to-one constraint on the abundance matrix; $\delta$ is the weight of the sum-to-one constraint; and $J_{M \times N}$ is the all-one matrix of size $M \times N$. The Equation (3-8) is usually solved by an alternative least square (ALS) minimization algorithm. The ALS algorithm iteratively updates the three vectors ($A_r, B_r, and\ c_r$) in sequence using a multiplicative update rule. Figure 3-3 shows the flowchart of the MVNTF. The multiplicative update rules adopted in MVNTF are listed below:

$$A \leftarrow A.* \frac{(X_{NK \times M}^T (B \odot C) + \delta J_{M \times N} B)}{(A(B \odot C)^T (B \odot C) + \delta AB^T B)},$$

$$(3\text{-}9)$$

$$A \leftarrow A.* \frac{(X_{NK \times M}^T (B \odot C) + \delta J_{M \times N} B)}{(A(B \odot C)^T (B \odot C) + \delta AB^T B)},$$

$$(3\text{-}10)$$

$$C \leftarrow C.* (X_{MN \times K}^T S)./(CS^T S),$$

$$(3\text{-}11)$$

$$S = [(A_1 \odot B_1) J_L \cdots (A_R \odot B_R) J_L],$$

$$(3\text{-}12)$$

where $(X_{NK \times M})_{(n-1)K+k,m} = x_{mnk}$ , $(X_{KM \times N})_{(k-1)M+m,n} = x_{mnk}$ , and $(X_{MN \times K})_{(m-1)N+n,k} = x_{mnk}$ are the unfolded tensor $\mathcal{X}$; and $J_L$ is an all-one column vector with length $L$.
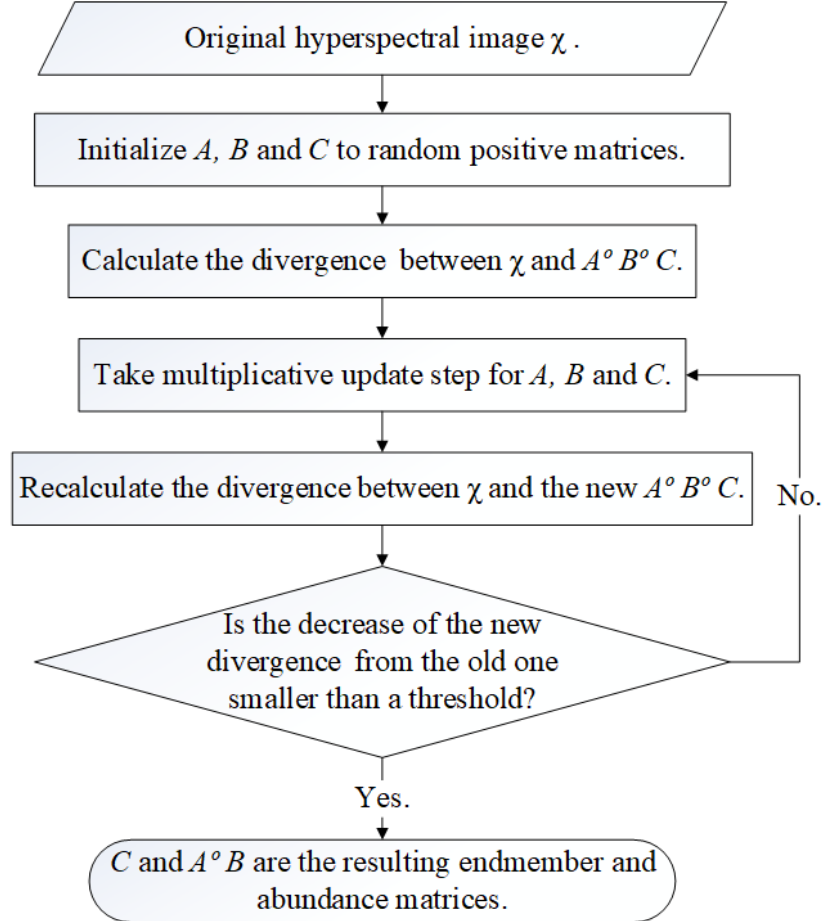


**Figure 3-3: Matrix vector nonnegative tensor factorization algorithm flowchart.**

### 3.3.4 Additional constraints

Due to the limited constraints, the MVNTF method may suffer from inefficient convergence. This problem also happens in NMF methods and various constraints have been added to the classical NMF to further confine the solution space. Three widely used constraints are sparseness, minimum volume, and outliner term. It is also promising to adopt them into the MVNTF framework. The MVNTF with the sparseness constraint is

referred to as sMVNTF from now on. One of the commonly agreed properties of the abundance matrix is the sparseness. It is often found that most of the pixels are mixtures of only a few of the endmembers. To implement this feature of the abundance matrix, this research added the norm of abundance matrix to the original cost function (Equation (3-8)):

$$min_{E,c} \parallel \mathcal{X} - \sum_{r=1}^{R} E_r \circ c_r \parallel_F^2 + \delta \parallel \sum_{r=1}^{R} E_r - J_{M \times N} \parallel_F^2 + \alpha \parallel \sum_{r=1}^{R} E_{r\parallel} \parallel_F^2,$$
$$s.t. A_r, B_r, c_r \geqslant 0,$$
(3-12)

where $\alpha$ is the sparseness parameter. This study tested eight sparseness parameters (0.01, 0.1, 0.2, 0.5, 0.7, 1, 5, and 10) in order to test the sensitivity of the constrained MVNTF on sparseness.

As mentioned in Chapter 2, the NMF-based SMA method can search outside the original image for endmembers. The MVNTF-based SMA methods share this advantage. However, even if the original image does not have any pure pixels, the endmembers should not locate too far away from the data cloud. The minimum volume constraint for endmembers addresses this problem by pulling the endmembers towards the data center. The basic idea is to minimize the simplex volume determined by the endmembers among all possible simplexes that circumscribe the data scatter space. This study adopts an approach of minimizing the sum of squared distances from the vertices to the centroid. The MVNTF with the minimum volume constraint is referred to as vMVNTF from now on. The cost function has correspondingly changed to:

$$min_{E,c} \parallel \mathcal{X} - \sum_{r=1}^{R} E_r \circ c_r \parallel_F^2 + \delta \parallel \sum_{r=1}^{R} E_r - J_{M \times N} \parallel_F^2 + \beta \parallel V \parallel_F^2,$$
$$V = \sum_{r=1}^{R} \parallel c_r - \frac{1}{R} \sum_{r'=1}^{R} c_{r'} \parallel_F^2,$$
(3-13)

$$s.t. A_r, B_r, c_r \geqslant 0,$$

where $\beta$ is the distance parameter; $\frac{1}{R} \sum_{r'=1}^{R} c_{r'}$ denotes the center of the current selected endmembers. We tried eight distance parameters: 0.01, 0.1, 0.2, 0.5, 0.7, 1, 5, and 10, in order to provide some insights.

In the end, due to the complex scattering effects that are often found in hyperspectral images, the spectral mixture may not be linear sometimes. A nonlinear term can be added in to the MVNTF model to account for such situation. The MVNTF with the nonlinear constraint is referred to as rMVNTF from now on. The reconstruction Equation (3-7) can be rewritten as below

$$\mathcal{X} = \sum_{r=1}^{R} E_r \circ c_r + N, \tag{3-14}$$

where $N$ represents the nonlinear term. Then the cost function is changed to:

$$min_{E,c} \parallel \mathcal{X} - \sum_{r=1}^{R} E_r \circ c_r \parallel_F^2 + \delta \parallel \sum_{r=1}^{R} E_r - J_{M \times N} \parallel_F^2 + \gamma \parallel N \parallel_F^2,$$
$$s.t. A_r, B_r, c_r \geqslant 0, \tag{3-15}$$

where $\gamma$ is the nonlinearity parameter. We tried eight more nonlinearity parameters: 0.01, 0.1, 0.2, 0.5, 0.7, 1, 5, and 10. The update rule for $N$ is derived based on the nonlinear term in NMF:

$$N_r \to N_r \frac{\sum_{r=1}^{R} \mathcal{X} / (E_r \circ c_r)}{c_r E_r^{-1} + \gamma \frac{N_r}{\parallel N_r \parallel_2}}. \tag{3-16}$$

## 3.4 Experiments and results

### 3.4.1 Study areas, reference spectra, and research design

Chapter 3 closely relates to the previous Chapter 2, as they share the same types of results of endmember matrices, impervious surface area (ISA) abundance maps, and ISA classification maps. Chapter 3 can be seen as an improvement from the method in Chapter 2, when the 2D matrix is replaced to the 3D tensor. To ensure the consistency and comparison between the two chapters, their studied images, reference spectra, and the overall research design are intentionally kept similar. This paper also used the two hyperspectral images (EnMAP and Hydice) introduced in 2.5.1. The simulated EnMAP image covers the whole Surrey, BC area and has patterns of urban and urban-suburban gradient environments (Figure 2-4). The Hydice image covers an urban area located in

Copperas Cove, TX (Figure 2-5). This paper used the same Hydice urban image as Chapter 2. The original simulated EnMAP image is not in a rectangle shape and has areas of no-data at corners. This irregular shape does not affect the NMF-based methods, as they perform on matricized hyperspectral images. But the MVNTF-based methods treat the hyperspectral image as a third-order tensor and cannot take data with no-data area. Thus, in this paper, the original simulated EnMAP image is cropped into a $300 \times 300$ pixels square shape, shown in Figure 3-4. This part of the EnMAP image also contains the urban and urban-suburban gradient areas. When comparing with the results from Chapter 2, the results from Chapter 2 were cropped into the same $300 \times 300$ square area.
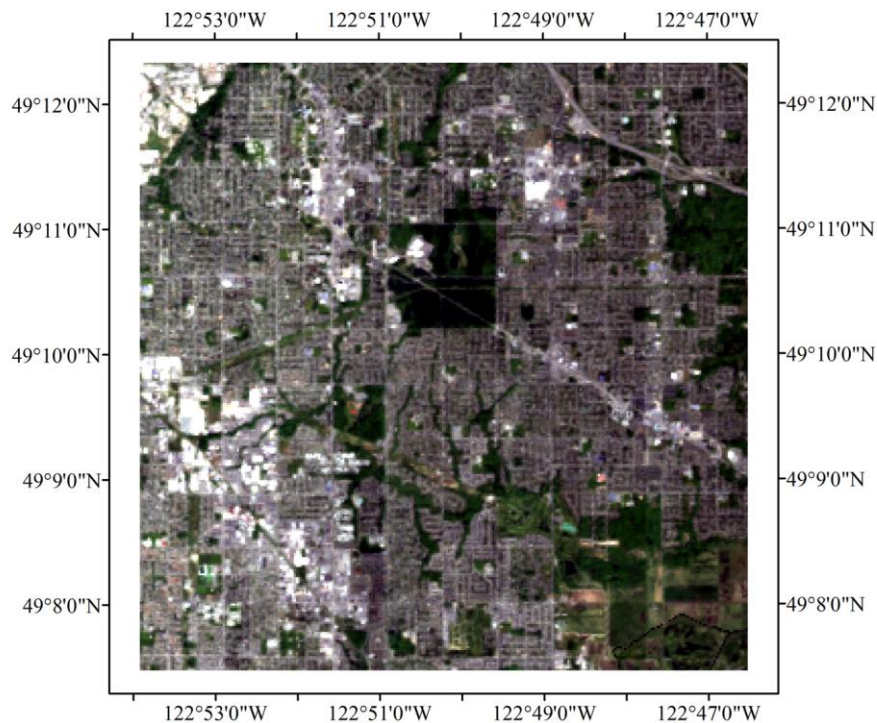


**Figure 3-4: The cropped simulated EnMAP image in RGB.**

The reference spectra that are used for identifying the extracted endmembers in this paper are the same sets produced in Chapter 2. In the case of the EnMAP image, the 20 endmembers of 18 different landuse/land cover (LULC) types were visually identified in the CASI images, the high-spatial-resolution airborne hyperspectral images that are used for simulating the EnMAP images (Figure 2-5). They are resampled to the spectral

instrument response functions for the EnMAP imager. The 18 LULC types consist of eight vegetation, nine impervious surface, and one soil. The Hydice urban image has the reference spectra of five: asphalt road, grass, tree, roof, and soil (Figure 2-6), available online. Another important component of the a priori knowledge for any SMA is the number of wanted endmembers. The wanted endmember numbers for the EnMAP and Hydice images are separately 27 and 5, which are decided by the Harsanyi–Farrand–Chang method (HFC). Detailed explanations concerning the determination of the number of wanted endmembers are stated in 2.4.1 Hyperspectral subspace identification.

This paper separately performed the MVNTF, sMVNTF, vMVNTF, and rMVNTF on the two studied images. Different column numbers $L$ for matrices $A_r$ and $B_r$ were tested and I found that using 90 as the column number provides the best results. Multiple sum-to-one constraint parameters $\delta$ were tested. I found that using 0.55 and 0.4 as the sum-to-one constraint parameter provides the best results correspondingly for the simulated EnMAP image and the Hydice urban image. The three constrained MVNTF were performed multiple times using eight different coefficients (0.01, 0.1, 0.2, 0.5, 0.7, 1, 5, and 10) for their constraints. In total, the MVNTF, sMVNTF, vMVNTF, and rMVNTF SMA methods with different constraint parameters on the added constraints composed 25 experiments on each of the studied image. Combining the method and the used constraint coefficient, the 25 experiments are referred to as MVNTF, sMVNTF (0.01), sMVNTF (0.2), sMVNTF (0.5), sMVNTF (0.7), sMVNTF (1), sMVNTF (5), sMVNTF (10), vMVNTF(0.01), vMVNTF (0.1), vMVNTF (0.2), vMVNTF (0.5), vMVNTF (0.7), vMVNTF (1), vMVNTF (5), vMVNTF (10), rMVNTF (0.01), rMVNTF (0.1), rMVNTF (0.2), rMVNTF (0.5), rMVNTF (0.7), rMVNTF (1), rMVNTF (5), and rMVNTF (10), where the value in the parenthesis is the used constraint coefficients. The reasons why the constraint coefficients determined by Equation (2-11) for the sparseness constraint and by Equation (2-15) for the nonlinearity are not included are: (1) the intrinsic mechanisms between the NMF- and MVNTF-based methods are different; and (2) the results from Equations (2-15) and Equation (2-15) in Chapter 2 show no advantages over the other constraints coefficients. The resulting endmember and abundance matrices were separately evaluated. The endmembers were first compared to the reference spectra and

identified using the spectral angle distance (SAD). To evaluate the abundance matrices, the same three indexes: the Pearson correlation coefficient of the linear regression model between the reference and predicted ISA percentages, mean absolute difference (MAD) between the reference and predicted ISA percentages, and the overall accuracy of the dominant abundance classification that were used in Chapter 2 were used in this paper. Detailed explanations concerning how the SMA results are evaluated can be found in 2.4.3 Evaluations. The following sections provide the resulting endmember matrices, ISA abundance maps, and ISA classification maps.

## 3.4.2    Extracted endmember

For the cropped simulated EnMAP image, the derived 25 sets of estimated endmembers were separately compared with the 20 resampled reference spectra, resulting in 25 SAD matrices. Some reference spectra that have similar spectral shapes (e.g. reference soil spectrum and several impervious reference spectra: red roof, asphalt and concrete), which makes it risky to assign the land cover type based on one SAD. The solution is explained in detailed in 2.5.3. Briefly, the reference spectra were first grouped into nine groups using the extended V-I-S models (Table 2-2). Then, the average SAD within each group was used to assign group type to the estimated endmembers from the SMA. In the results, the majority of the tested MVNTF, sMVNTF, vMVNTF, and rMVNTF methods can successfully extract both pervious and impervious endmembers, where pervious endmembers often occupy more than half of the positions among the total obtained endmember sets. Some constrained MVNTF with certain coefficients failed to extract any of the impervious surface and they are sMVNTF (5), sMVNTF (10), vMVNTF (5), vMVNTF (10), and rMVNTF (10). For the Hydice urban image, the derived 25 sets of estimated endmembers were separately compared with the five reference spectra. In the resulting SAD matrices, we found that, in Chapter 2 the single SAD value between each estimated endmember and each reference endmember can successfully identify the land cover types, without any grouping process. As with the EnMAP image, the majority of the tested MVNTF, sMVNTF, vMVNTF, and rMVNTF methods can successfully extract both pervious and impervious endmembers from the Hydice urban image. The pervious

endmembers also often occupy more than half of the positions among the total obtained endmember sets. The rMVNTF (10) is the only experiment that did not find any impervious endmembers.

The endmember SAD results are in Table 3-2 for the simulated EnMAP image and in Table 3-3 for the Hydice image. The minimum and median SAD values and their differences within each SAD matrix are shown. As mentioned before, good endmembers with significant spectral signature should be close to the target reference spectrum and be far from other reference spectra, which means smaller minimum value, larger median value, and larger differences between the two. For the constrained NMF methods, only the SAD results with the maximum differences between the minimum and median SAD values are shown.

**Table 3-1: Average minimum/median/difference spectral angle distance values (EnMAP).**

| Average SAD | MVNTF | sMVNTF (0.01) | vMVNTF (0.01) | rMVNTF (0.01) |
|---|---|---|---|---|
| Minimum | 0.34 | 0.34 | 0.34 | 0.34 |
| Median | 0.71 | 0.71 | 0.71 | 0.71 |
| Difference | 0.37 | 0.37 | 0.37 | 0.37 |

**Table 3-2: Average minimum/median/difference spectral angle distance values (Hydice).**

| Average SAD | MVNTF | sMVNTF (0.7) | vMVNTF (1) | rMVNTF (1) |
|---|---|---|---|---|
| Minimum | 0.37 | 0.30 | 0.30 | 0.31 |
| Median | 0.77 | 0.79 | 0.79 | 0.79 |
| Difference | 0.40 | 0.49 | 0.49 | 0.48 |

## 3.4.3    ISA abundances and classification maps

As in Chapter 2, we grouped the abundance matrices according to the identified endmembers. There are 25 resulting ISA abundance maps for each of the studied images. For the simulated EnMAP image, the reference impervious abundance map was derived from the original CASI images. For the Hydice urban image, the reference abundance

map is available online (Zhu et al. 2014). A set of 500 pair-samples were randomly selected from each predicted map and the reference ISA abundance maps. The sets of samples are used to produce linear regression models between the predicted and reference ISA abundances. The better the fitted linear regression model the closer the predicted ISA abundance values to the reference ones. The model parameters are listed in Table 3-3 (EnMAP) and 3-4 (Hydice), where the value in the parenthesis after the SMA method is the used constraint parameter. The abundance maps obtained from the MVNTF-based method also suffer from an underestimation of ISA abundance. Thus, normalizations have been performed. Although great consistency can be found between the predicted and reference ISA abundance maps from the linear regression results in Table 3-3 and 3-4, the predicted ISA abundance maps severely underestimate the ISA area. The same reasoning in 2.3.4 applies here. The same procedure in Chapter 2 has been adopted here to convert the predicted ISA abundance maps to ISA binary classification maps. The last columns in Table 3-3 and 3-4 show the overall accuracy for the classification results.

In the case of the EnMAP image, the correlation coefficients range from 0.75 to 0.79. Considering the 25 processed MVNTF-based experiments, the traditional MVNTF method achieves the lowest correlation coefficients of 0.75. Except for certain constraint coefficients (sMVNTF (5), sMVNTF (10), vMVNTF (5), vMVNTF (10), and rMVNTF (10)), the three added constraints at different level increased the linear regression accuracy. The sMVNTF (0.2) and vMVNTF (0.1) simultaneously achieved the highest linear regression correlation coefficients of 0.79. The results from rMVNTF show very small variations among different nonlinear coefficients with a universal correlation coefficient of 0.75. Regarding the MAD values, the sMVNTF (0.2) and vMVNTF (0.1) method also out performs the traditional MVNTF method and achieves the lowest values. The classification accuracies derived for the simulated EnMAP image reached 85%. The three constrained MVNTF methods all achieved better classification accuracies than the traditional MVNTF.

**Table 3-3: Reference and predicted ISA abundance linear regression parameters and classification overall accuracy (EnMAP).**

| Method | R | P-value | MAD | Accuracy |
|---|---|---|---|---|
| **MVNTF** | **0.75** | **9.0e-6** | **0.1738** | **83.4%** |
| sMVNTF (0.01) | 0.75 | 3.2e-5 | 0.2100 | 83.6% |
| sMVNTF (0.1) | 0.77 | 5.3e-5 | 0.2039 | 83.7% |
| **sMVNTF (0.2)** | **0.79** | **1.3e-7** | **0.1802** | **85.1%** |
| sMVNTF (0.5) | 0.75 | 5.6e-5 | 0.2671 | 84.6% |
| sMVNTF (0.7) | 0.75 | 7.8e-5 | 0.2683 | 84.8% |
| sMVNTF (1.0) | 0.76 | 1.9e-6 | 0.2714 | 84.8% |
| sMVNTF (5.0) | - | - | - | - |
| sMVNTF (10.0) | - | - | - | - |
| vMVNTF (0.01) | 0.75 | 5.8e-5 | 0.1996 | 83.7% |
| **vMVNTF (0.1)** | **0.79** | **9.1e-7** | **0.1856** | **84.8%** |
| vMVNTF (0.2) | 0.75 | 8.2e-5 | 0.2671 | 84.5% |
| vMVNTF (0.5) | 0.75 | 9.6e-4 | 0.2689 | 84.8% |
| vMVNTF (0.7) | 0.76 | 6.5e-5 | 0.2713 | 84.8% |
| vMVNTF (1.0) | 0.76 | 6.3e-5 | 0.2739 | 84.7% |
| vMVNTF (5.0) | - | - | - | - |
| vMVNTF (10.0) | - | - | - | - |
| **rMVNTF (0.01)** | **0.75** | **2.3e-5** | **0.1988** | **83.5%** |
| rMVNTF (0.1) | 0.75 | 3.6e-5 | 0.1988 | 83.4% |
| rMVNTF (0.2) | 0.75 | 9.8e-6 | 0.1988 | 83.4% |
| rMVNTF (0.5) | 0.75 | 5.1e-5 | 0.1989 | 83.4% |
| rMVNTF (0.7) | 0.75 | 3.5e-7 | 0.1990 | 83.5% |
| rMVNTF (1.0) | 0.75 | 5.3e-6 | 0.1990 | 83.5% |
| rMVNTF (5.0) | 0.75 | 4.6e-5 | 0.1989 | 83.4% |
| rMVNTF (10.0) | - | - | - | - |

The linear regression results of Hydice urban image also show an underestimation in the ISA abundance, where the slopes are around 0.19-0.24 and the y-intercepts are between 0.11-0.23. The correlation coefficients from MVNTF-based methods have a larger range

(0.58 to 0.75) than in the simulated EnMAP case. The traditional MVNTF methods achieve a moderate linear regression correlation coefficient of 0.72. With the sparseness, volume, and nonlinearity constraints, the highest achieved linear regression correlation coefficients are separately 0.75, 0.75, and 0.74. The constraint parameter values used for the highest linear regression accuracy are 0.2 for sparseness, 0.5 for volume, and 0.2/0.5/0.7/5 for nonlinearity. The three constrained MVNTF methods also achieve smaller MAD values than the traditional MVNTF method when using the appropriate constraint parameters. The classification accuracies derived for the Hydice urban image reached 85%. The three constrained MVNTF methods all achieved better classification accuracies than the traditional MVNTF.

**Table 3-4: Reference and predicted ISA abundance linear regression parameters and classification overall accuracy (Hydice).**

| Method | R | P-value | MAD | Accuracy |
|---|---|---|---|---|
| **MVNTF** | **0.72** | **2.8e-6** | **0.2297** | **85.1%** |
| sMVNTF (0.01) | 0.72 | 1.8e-5 | 0.2297 | 85.2% |
| sMVNTF (0.1) | 0.73 | 2.5e-5 | 0.2292 | 86.1% |
| **sMVNTF (0.2)** | **0.75** | **4.3e-6** | **0.2185** | **87.0%** |
| sMVNTF (0.5) | 0.74 | 6.3e-6 | 0.2339 | 86.8% |
| sMVNTF (0.7) | 0.73 | 8.1e-6 | 0.2357 | 86.5% |
| sMVNTF (1.0) | 0.73 | 5.6e-5 | 0.2378 | 84.9% |
| sMVNTF (5.0) | 0.58 | 1.2e-3 | 0.2592 | 79.2% |
| sMVNTF (10.0) | 0.62 | 3.5e-5 | 0.2540 | 81.6% |
| vMVNTF (0.01) | 0.72 | 7.5e-5 | 0.2297 | 85.4% |
| vMVNTF (0.1) | 0.73 | 1.3e-6 | 0.2294 | 85.8% |
| vMVNTF (0.2) | 0.74 | 5.4e-5 | 0.2292 | 86.5% |
| **vMVNTF (0.5)** | **0.75** | **6.9e-6** | **0.2315** | **87.0%** |
| vMVNTF (0.7) | 0.74 | 2.8e-5 | 0.2330 | 86.9% |
| vMVNTF (1.0) | 0.73 | 9.2e-5 | 0.2348 | 85.0% |
| vMVNTF (5.0) | 0.66 | 4.1e-5 | 0.2493 | 84.0% |
| vMVNTF (10.0) | - | - | - | - |
| rMVNTF (0.01) | 0.72 | 5.2e-5 | 0.2297 | 85.1% |
| rMVNTF (0.1) | 0.73 | 5.4e-5 | 0.2291 | 85.5% |
| **rMVNTF (0.2)** | **0.74** | **6.1e-5** | **0.2292** | **86.4%** |
| rMVNTF (0.5) | 0.74 | 1.6e-5 | 0.2320 | 86.3% |
| rMVNTF (0.7) | 0.74 | 2.5e-5 | 0.2338 | 85.7% |
| rMVNTF (1.0) | 0.73 | 3.3e-5 | 0.2361 | 85.2% |
| rMVNTF (5.0) | 0.74 | 3.6e-5 | 0.2325 | 85.5% |
| rMVNTF (10.0) | - | - | - | - |

## 3.5 Discussion

Using four MVNTF (matrix vector nonnegative tensor factorization) -based (MVNTF, sMVNTF, vMVNTF, and rMVNTF) SMA methods, 25 sets of endmembers, ISA abundance and classification maps were generated for each of the studied images (the simulated EnMAP and Hydice urban images). The results proved to have great potential

for the future use of MVNTF-based SMA methods. Specifically, what we learned from these results can be discussed in two parts: (1) the comparison of the ISA abundance maps among different SMA methods and between the two study images; (2) the comparison of the ISA classification maps among different SMA methods and between the two studied images.

## 3.5.1    ISA abundance maps

The resulting ISA abundance maps are shown in Figure 3-5 for the simulated EnMAP image and in Figure 3-6 for the Hydice urban image. In the first row of Figure 3-5, the reference ISA abundance map is shown at left, and the MVNTF-derived ISA abundance map is shown at right. In the second row of Figure 3-5, only three constrained MVNTF ISA abundance maps were shown for each image, corresponding to the sMVNTF, vMVNTF, and rMVNTF results with the best correlation coefficients. The traditional MVNTF and the three constrained MVNTF methods shared two common problems: overestimate the medium ISA abundance values (circle A) and underestimate the low ISA abundance values (circle B). The constrained MVNTF methods suffer less from these two problems.
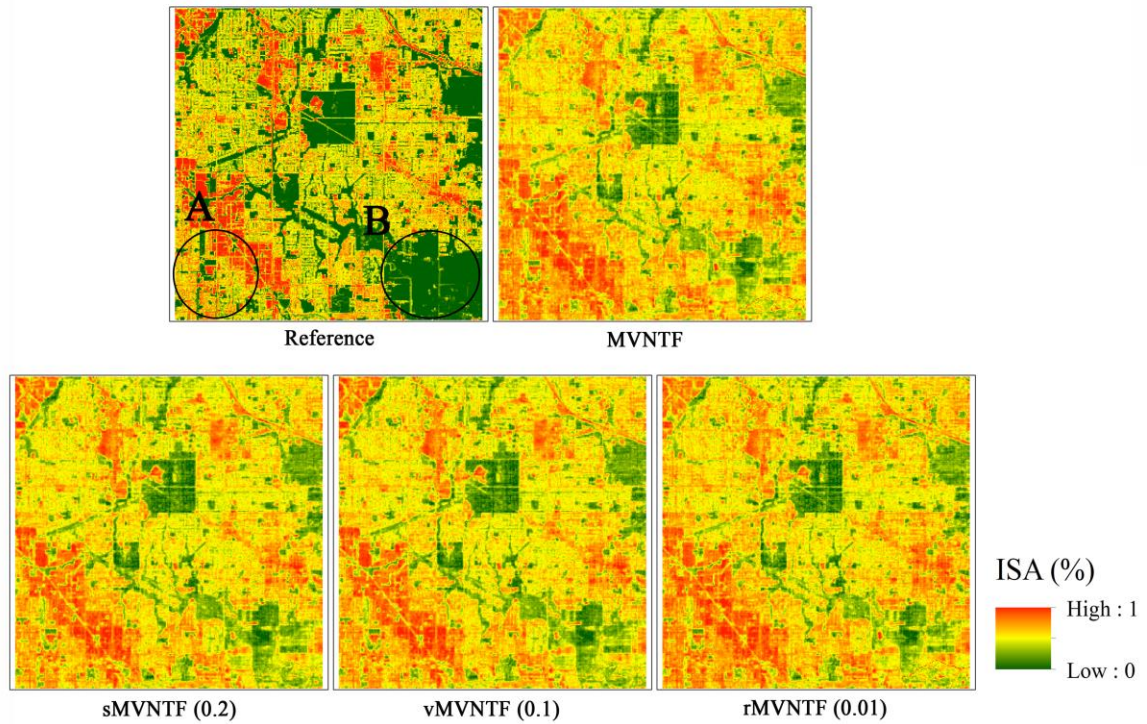
**Figure 3-5: The best impervious surface area abundance maps from each spectral mixture analysis method (EnMAP).**
**(the value in the parenthesis is the used constraint parameter)**

The Hydice urban image ISA abundance results are shown in Figure 3-6. In the first row of Figure 3-6, the reference ISA abundance map is shown at left, and the MVNTF-derived ISA abundance map is shown at right, which shows consistency with the reference ISA abundance map. In the second row of Figure 3-6, only three constrained MVNTF ISA abundance maps were shown for each image, corresponding to the sMVNTF, vMVNTF, and rMVNTF results with the best correlation coefficients. In the medium and high ISA abundance areas, the traditional MVNTF and the three constrained MVNTF methods show similar detection ability. In low ISA abundance areas, the traditional MVNTF method greatly overestimate the ISA abundance, while the three constrained MVNTF methods better depicted these areas (circle A and B).
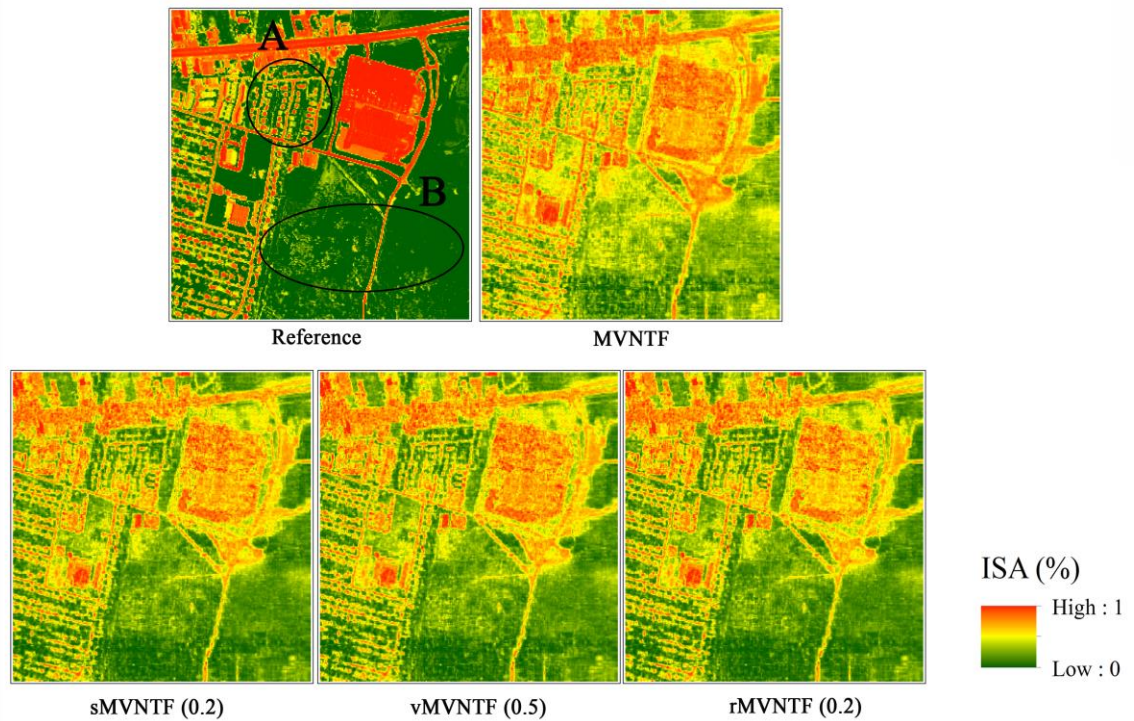
**Figure 3-6: The best impervious surface area abundance maps from each spectral mixture analysis method (Hydice).**
**(the value in the parenthesis is the used constraint parameter)**

## 3.5.2    ISA classification maps

The classification results are shown in Figure 3-7 for the simulated EnMAP image and in Figure 3-8 for the Hydice urban image. Begin with the results from the simulated EnMAP image. In the first row of Figure 3-7, the reference ISA classification map is shown at left.  On the right side of Figure 3-7, the result from the MVNTF appears to be not sensitive to small ISA, as narrow roads and small roofs are often misclassified as pervious area. In the second row of Figure 3-7, only three constrained MVNTF ISA abundance maps were shown for each image, corresponding to the sMVNTF, vMVNTF, and rMVNTF results with the best classification overall accuracy. The three constrained MVNTF methods shared the same insensitivity towards the small ISA area (circle A).
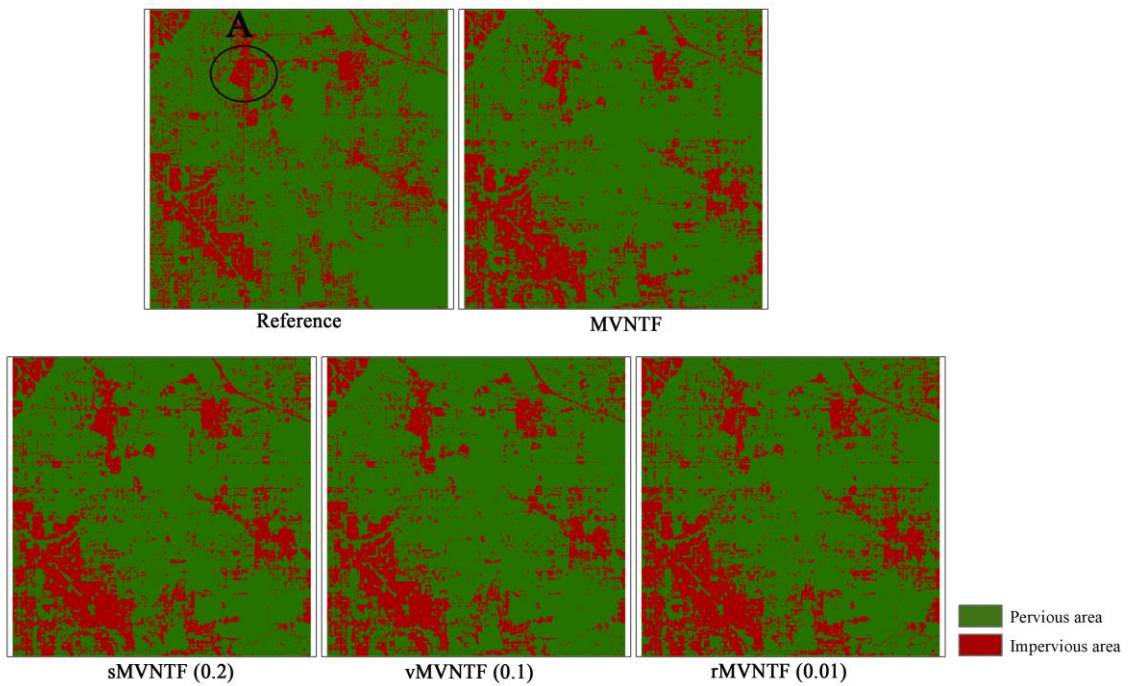
**Figure 3-7: The best impervious surface area classification maps from each spectral mixture analysis method (EnMAP).**
**(the value in the parenthesis is the used constraint parameter)**

In the case of the Hydice urban image, Figure 3-8 shows the reference ISA classification map at the left, and the MVNTF-derived ISA abundance map is shown at right. The MVNTF method misclassified parts of the parking lot and bright roof in the upper-right corner as pervious area (circle A). In addition, some parts of the soil in the middle right part of the image is misclassified as impervious (circle B). In the second row of Figure 3-8, only three constrained MVNTF ISA abundance maps were shown for each image, corresponding to the sMVNTF, vMVNTF, and rMVNTF results with the best classification overall accuracy. At less significant levels than the traditional MVNTF method, the three constrained MVNTF results also misclassified the upper right parking lot or the bright roof next to it as pervious; and misclassified the lower right vegetation as impervious.
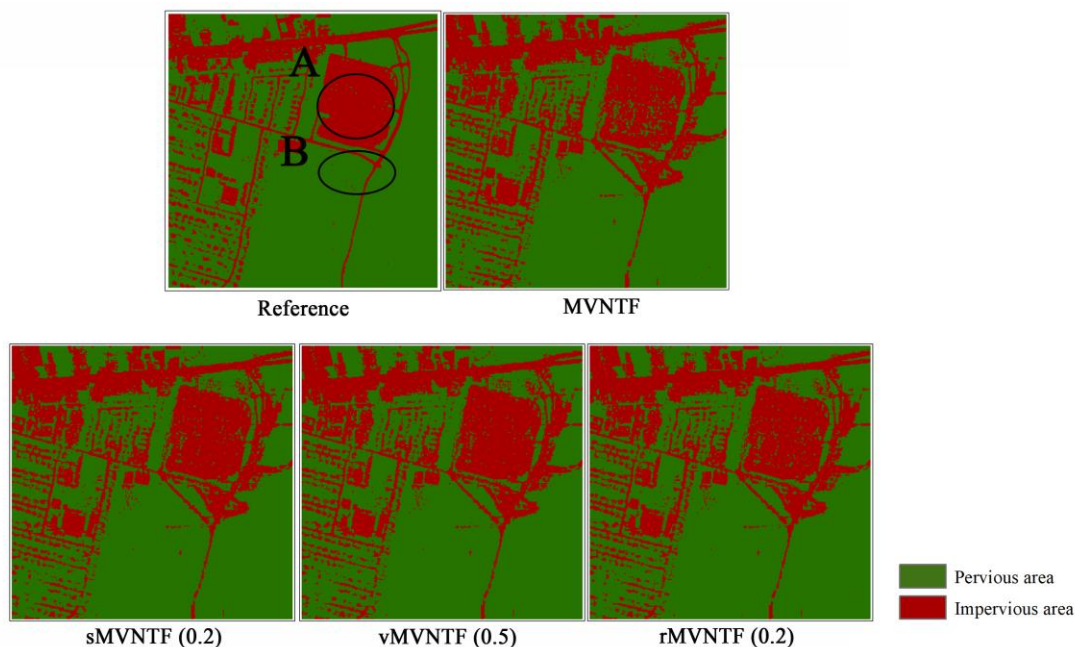
**Figure 3-8: The best impervious surface area abundance maps from each spectral mixture analysis method (Hydice).**
**(the value in the parenthesis is the used constraint parameter)**

## 3.5.3 Comparison between the matrix-based and tensor-based methods

By using the same data sources as Chapter 2, the tensor-based spectral unmixing results from this Chapter can be compared to the matrix-based spectral unmixing results. In the case of the simulated EnMAP image, because of the limitation of the tensor-based methods on the shape of the input data, the original data was cropped. Thus, in order to make accurate comparison, I cropped the EnMAP results from Chapter 2 to be in the same location. The conclusion between Chapter 2 and Chapter 3 results can be found in Table 3-5 for the EnMAP image and in Table 3-6 for the Hydice image. Table 3-4 shows that the tensor-based methods outperform the matrix-based methods for the simulated EnMAP image. Yet, for the Hydice image, no clear advantages can be found in the tensor-based methods over the matrix-based methods. The reason for the different comparison results may probably lie in the differences in the two datasets. The simulated EnMAP image (30m) has a much coarser spatial resolution than the Hydice image (2m). Thus, it is possible that the tensor-based subpixel methods work better with the medium spatial resolution images than with the high spatial resolutions images. The possible

reason may be that the medium-spatial-resolution image has a smoother spatial texture, which can be easily learned by the tensor-representation, while the high-spatial-resolution image has a heterogeneous spatial texture, which causes confusion between intra- and inter-class items.

**Table 3-5: Comparison between NMF-based and MVNTF-based methods (EnMAP).**

| Method | R Matrix-based | R Tensor-based | MAD Matrix-based | MAD Tensor-based | Accuracy Matrix-based | Accuracy Tensor-based |
|---|---|---|---|---|---|---|
| Traditional | 0.71 | 0.75 | 0.2128 | 0.1989 | 82.6% | 83.4% |
| Sparseness (average) | 0.68 | 0.76 | 0.2511 | 0.2317 | 81.0% | 84.4% |
| Sparseness (lowest) | 0.64 | 0.75 | 0.2125 | 0.1802 | 77.6% | 83.6% |
| Sparseness (highest) | 0.72 | 0.79 | 0.2848 | 0.2714 | 83.6% | 84.8% |
| Volume (average) | 0.70 | 0.76 | 0.2760 | 0.2444 | 82.5% | 84.5% |
| Volume (lowest) | 0.69 | 0.75 | 0.2122 | 0.1856 | 80.3% | 83.7% |
| Volume (highest) | 0.73 | 0.79 | 0.2900 | 0.2739 | 84.6% | 84.8% |
| Nonlinear (average) | 0.70 | 0.75 | 0.2791 | 0.1989 | 82.1% | 83.4% |
| Nonlinear (lowest) | 0.68 | 0.75 | 0.2449 | 0.1988 | 79.75% | 83.4% |
| Nonlinear (highest) | 0.72 | 0.75 | 0.2996 | 0.1990 | 82.5% | 83.5% |

**Table 3-6: Comparison between NMF-based and MVNTF-based methods (Hydice).**

| Method | R Matrix-based | R Tensor-based | MAD Matrix-based | MAD Tensor-based | Accuracy Matrix-based | Accuracy Tensor-based |
|---|---|---|---|---|---|---|
| **Traditional** | 0.70 | 0.72 | 0.2004 | 0.2297 | 87.0% | 85.10% |
| **Sparseness (average)** | 0.68 | 0.71 | 0.2294 | 0.2373 | 84.1% | 84.7% |
| **Sparseness (lowest)** | 0.46 | 0.58 | 0.1590 | 0.2185 | 75.5% | 79.2% |
| **Sparseness (highest)** | 0.83 | 0.75 | 0.3306 | 0.2592 | 91.8% | 87.0% |
| **Volume (average)** | 0.64 | 0.73 | 0.2112 | 0.2338 | 82.2% | 85.8% |
| **Volume (lowest)** | 0.32 | 0.66 | 0.1693 | 0.2292 | 72.8% | 84.0% |
| **Volume (highest)** | 0.80 | 0.75 | 0.3189 | 0.2493 | 87.9% | 87.0% |
| **Nonlinear (average)** | 0.62 | 0.73 | 0.2142 | 0.2318 | 83.0% | 85.7% |
| **Nonlinear (lowest)** | 0.42 | 0.72 | 0.2004 | 0.2291 | 77.6% | 85.1% |
| **Nonlinear (highest)** | 0.71 | 0.74 | 0.2338 | 0.2361 | 87.5% | 86.4% |

## 3.5.4    Computation time

Lastly, the processing times in seconds for the MVNTF-based hyperspectral SMA methods are shown in Table 3-7. There are separately 90000 and 94249 pixels in the simulated EnMAP image and the Hydice urban image that need to be processed. It can be observed that for both the studied images the three added constraints helped to reduce the processing time. This reduction in time is probably due to the narrowed solution space by the constraints. In the cases of the two-studied images, one common trend occurs in all the three constraints that larger coefficients lead to better time efficiency, as larger constraints exert more force on the solution. Despite that the sizes of the two studied images are very similar when the Hydice urban image is slightly larger than the EnMAP image, the processing times of the EnMAP image are much longer than those of the Hydice urban image. Thus, it can be deduced that the size of an image plays a small role in determining the processing time of the MVNTF-based unmixing, while the internal features of an image determine the time to find the solution.

**Table 3-7: Processing times in seconds.**

| Method | Time (EnMAP) | Time (Hydice) |
|---|---|---|
| MVNTF | 4810.3 | 834.8 |
| sMVNTF (0.01) | 5731.2 | 842.1 |
| sMVNTF (0.1) | 3934.5 | 874.6 |
| sMVNTF (0.2) | 3741.1 | 734.1 |
| sMVNTF (0.5) | 2458.2 | 634.6 |
| sMVNTF (0.7) | 2417.1 | 581.2 |
| sMVNTF (1.0) | 2003.6 | 525.4 |
| sMVNTF (5.0) | - | 219.3 |
| sMVNTF (10.0) | - | 265.7 |
| vMVNTF (0.01) | 2458.2 | 848.0 |
| vMVNTF (0.1) | 3170.8 | 882.1 |
| vMVNTF (0.2) | 3222.1 | 795.7 |
| vMVNTF (0.5) | 2425.9 | 697.5 |
| vMVNTF (0.7) | 2160.9 | 643.6 |
| vMVNTF (1.0) | 1801.2 | 582.9 |
| vMVNTF (5.0) | - | 307.9 |
| vMVNTF (10.0) | - | 262.7 |
| rMVNTF (0.01) | 5616.0 | 1061.8 |
| rMVNTF (0.1) | 5394.8 | 991.6 |
| rMVNTF (0.2) | 5298.3 | 891.9 |
| rMVNTF (0.5) | 5142.6 | 710.4 |
| rMVNTF (0.7) | 5312.2 | 611.3 |
| rMVNTF (1.0) | 4851.9 | 566.3 |
| rMVNTF (5.0) | 4788.4 | 641.1 |
| rMVNTF (10.0) | - | - |

## 3.6  Conclusion

This research adopts the newly proposed MVNTF unmixing method for the hyperspectral SMA problem. Three additional constraints (sparseness, volume, and nonlinearity) were added to the plain MVNTF method, in order to improve the accuracy of impervious mapping. Two hyperspectral images with different spectral/spatial resolution and in

different scenes were studied: the simulated EnMAP image of Surrey, BC, Canada and the Hydice urban image of Copperas Cove, TX, USA. Four MVNTF-based SMA methods were tested: MVNTF, sMVNTF, vMVNTF, and rMVNTF. We analyzed the ISA mapping results from the aspects of different images and different SMA methods. The three research questions in 3.2 were correspondingly addressed below.

(1) We found that the MVNTF-based hyperspectral SMA methods can successfully derive ISA abundance and classification maps. The comparison between the tensor-based spectral unmixing from Chapter 3 and the matrix-based spectral unmixing from Chapter 2 shows that the tensor-based methods obtained a higher accuracy with the simulated EnMAP image than the Hydice urban image.

(2) By equipping three different constraints for SMA purposes, the constrained MVNTF-based methods provide better accuracies in ISA abundance and classification maps for both of the studied images. Using the constrained MVNTF methods, the simulated EnMAP image achieves correlation coefficients of 0.79, compared to the correlation coefficient of 0.75 from the traditional MVNTF method. The Hydice urban image achieves correlation coefficients of 0.75, compared to the correlation coefficient of 0.72 from the traditional MVNTF. From the abundance maps, the pervious/impervious classification maps were generated using thresholds. For both images, the constrained MVNTF methods improved the overall accuracy by 2% from the traditional MVNTF and reached ~85%, which are as good as the up-to-date ISA studies (Zhang, Weng and Shao 2017, Zhang and Weng 2016), which often obtained help from the additional spectral indices and multi-date or multi-sensor data.

(3) The abundance maps directly obtained from the MVNTF-based methods underestimated the ISA abundance greatly. It is necessary to perform a linear normalization on the obtained abundance maps to achieve reliable results.

(4) The tensor-based image processing methods have gradually gained attention in the remote sensing community. The ISA mapping is among the major topics of remote sensing studies. No attempt has been made to apply the tensor-based SMA in the ISA

mapping before. This study proved that the tensor-based SMA methods have some potential in future ISA mapping.

## References

Arngren, M., M. N. Schmidt & J. Larsen (2011) Unmixing of hyperspectral images using Bayesian non-negative matrix factorization with volume prior. *Journal of Signal Processing Systems,* 65**,** 479-496.

Bioucas-Dias, J. M., A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader & J. Chanussot (2012) Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE journal of selected topics in applied earth observations and remote sensing,* 5**,** 354-379.

De Lathauwer, L. (2008) Decompositions of a higher-order tensor in block terms—Part I: Lemmas for partitioned matrices. *SIAM Journal on Matrix Analysis and Applications,* 30**,** 1022-1032.

Deng, Y.-J., H.-C. Li, L. Pan, L.-Y. Shao, Q. Du & W. J. Emery (2018) Modified Tensor Locality Preserving Projection for Dimensionality Reduction of Hyperspectral Images. *IEEE Geoscience and Remote Sensing Letters*.

Du, B., M. Zhang, L. Zhang, R. Hu & D. Tao (2017) PLTD: Patch-based low-rank tensor decomposition for hyperspectral images. *IEEE Transactions on Multimedia,* 19**,** 67-79.

Fang, L., N. He & H. Lin (2017) CP tensor-based compression of hyperspectral images. *JOSA A,* 34**,** 252-258.

Févotte, C. & N. Dobigeon (2015) Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization. *IEEE Transactions on Image Processing,* 24**,** 4810-4819.

Gao, Y., X. Wang, Y. Cheng & Z. J. Wang (2015) Dimensionality reduction for hyperspectral data based on class-aware tensor neighborhood graph and patch alignment. *IEEE transactions on neural networks and learning systems,* 26**,** 1582-1593.

Kolda, T. G. & B. W. Bader (2009) Tensor decompositions and applications. *SIAM review,* 51**,** 455-500.

Liu, Z., B. Tang, X. He, Q. Qiu & H. Wang (2017) Sparse Tensor-Based Dimensionality Reduction for Hyperspectral Spectral–Spatial Discriminant Feature Extraction. *IEEE Geoscience and Remote Sensing Letters,* 14**,** 1775-1779.

Miao, L. & H. Qi (2007) Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. *IEEE Transactions on Geoscience and Remote Sensing,* 45**,** 765-777.

Qian, Y., F. Xiong, S. Zeng, J. Zhou & Y. Y. Tang (2017) Matrix-vector nonnegative tensor factorization for blind unmixing of hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing,* 55**,** 1776-1792.

Ren, Y., L. Liao, S. J. Maybank, Y. Zhang & X. Liu (2017) Hyperspectral image spectral-spatial feature extraction via tensor principal component analysis. *IEEE Geoscience and Remote Sensing Letters,* 14**,** 1431-1435.

Veganzones, M. A., J. E. Cohen, R. C. Farias, J. Chanussot & P. Comon (2016) Nonnegative tensor CP decomposition of hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing,* 54**,** 2577-2588.

Zhang, L. & Q. Weng (2016) Annual dynamics of impervious surface in the Pearl River Delta, China, from 1988 to 2013, using time series Landsat imagery. *ISPRS Journal of Photogrammetry and Remote Sensing,* 113**,** 86-96.

Zhang, L., Q. Weng & Z. Shao (2017) An evaluation of monthly impervious surface dynamics by fusing Landsat and MODIS time series in the Pearl River Delta, China, from 2000 to 2015. *Remote Sensing of Environment,* 201**,** 99-114.

Zhang, Q., H. Wang, R. Plemmons & V. P. Pauca. 2007. Spectral unmixing using nonnegative tensor factorization. In *Proceedings of the 45th annual southeast regional conference*, 531-532. ACM.

Zhang, Q., H. Wang, R. J. Plemmons & V. P. l. Pauca (2008) Tensor methods for hyperspectral data analysis: a space object material identification study. *JOSA A,* 25**,** 3001-3012.

Zhu, F., Y. Wang, S. Xiang, B. Fan & C. Pan (2014) Structured sparse method for hyperspectral unmixing. *ISPRS Journal of Photogrammetry and Remote Sensing,* 88**,** 101-118.

# 4 The application of spectral-spatial representation of hyperspectral image in dimension reduction

Progress in hyperspectral imagery collection and storage during the past decade provides new opportunities and challenges. Although hyperspectral imagery with high spectral resolution provides a wealth of information, the problem known as the *curse of dimensionality* is an obstacle in myriad hyperspectral imagery applications (Khodr and Younes 2011). The necessity of dimension reduction (DR) originates in the redundant and noise-abundant nature of hyperspectral data. An appropriate dimension reduction (DR) method prepares data for more effective information retrieval by revealing low-dimensional structures hidden in high-dimensional spaces. Currently, the majority of DR methods are spectral-based methods. They focus on preserving the useful information from the spectral-domain in the DR process, when the information regarding the spatial interpixel correlation is discarded. Recently, a genre of spectral-spatial DR methods were proposed to address this problem. Two different spectral-spatial strategies have been successively developed: patch-based (Zhou, Peng and Chen 2015, Pu et al. 2014), and tensor-patch-based methods (Velasco-Forero and Angulo 2013, Zhang et al. 2013, Deng et al. 2018). In detail, the patch-based DR methods learn the spatial context about a pixel, as they measure the similarity between two pixels using the two corresponding neighborhoods. The tensor-patch-based DR methods benefit from the recently rising multilinear algebra solution and can directly process the 3D hyperspectral data to preserve the spatial context about a pixel. Both as new spectral-spatial solutions in hyperspectral DR, the patch-based and tensor-patch-based methods have the potential to play important roles in future remote sensing studies.

## 4.1  Background

### 4.1.1  Traditional dimension reduction methods

The dimension reduction (DR) methods can be roughly divided into two categories:

unsupervised and supervised. This research focuses on the unsupervised DR methods. The traditional unsupervised DR methods, including principle component analysis (PCA) (Pearson 1901) and multidimensional scaling (MDS) (Torgerson 1952), are based on linear transformations. Specifically, the objective of the PCA algorithm is to maximize data variance in the new projection. Mathematically, this objective can be stated as finding a linear projection $M$ that maximizes the cost function $trace(M^T cov(X)M)$, where $cov(X)$ is the sample covariance matrix of the data $X$. This mathematical problem can be solved by the eigenproblem: (Hotelling 1933). On the other hand, the MDS aims to preserve the pairwise distances between data points during the projection. Mathematically, this objective can be stated as finding a linear projection $M$ that minimizes the cost function $\sum_{ij}(\| x_i - x_j \|^2 - \| y_i - y_j \|^2)$, where $\| x_i - x_j \|^2$ is the Euclidean distance between high-dimensional data points $x_i$ and $x_j$ and $\| y_i - y_j \|^2$ is the Euclidean distance between low-dimensional data points $y_i$ and $y_j$. This mathematical problem can be solved by the eigenproblem of the Gram matrix ($K = XX^T$) of the data $X$. The PCA and MDS methods have been successfully applied in many DR problems. However, with linear transformations, the PCA and MDS methods may not be adequate for complex nonlinear datasets. To address this issue, nonlinear variations of the PCA and MDS have been proposed (Lee and Verleysen 2007). The kernel PCA is a straightforward nonlinear version of PCA. The kernel method has been widely used to convert the linear techniques into nonlinear (Shawe-Taylor and Cristianini 2004). The kernel PCA reformulates the linear PCA in a high-dimensional space using a chosen kernel function. The kernel PCA solves the eigenproblem of the kernel matrix instead of the covariance matrix, which generates nonlinear projections (Schölkopf, Smola and Müller 1998). Several MDS nonlinear variations were proposed, including Isomap. The Isomap method substitutes the Euclidean distance in the MDS with geodesic distance (Tenenbaum, De Silva and Langford 2000), which can also be considered as a kernel function.

The cost functions of the PCA, MDS, and their nonlinear variations are solved based on a full matrix. Thus, they naturally weigh large data variance/pairwise distances over small ones (Van Der Maaten, Postma and Van den Herik 2009), which leads to better

preservation of global information than local structure. In hyperspectral feature extraction and classification, the local structure contributes to the differentiation between large intra-class variation and small inter-class variation, probably caused by high spatial resolution and the large scale of the image. Thus, the above global-information-preserving DR methods may not be optimal solutions for hyperspectral data (Cheriyadat and Bruce 2003). Fortunately, a group of graph-based DR methods have been proposed in the beginning of 2000 (Lee and Verleysen 2012). These methods learned the data local structure from adjacency graphs/weight matrices, which represents the topology information in the data. Based on geometrical theory, the local information can be embedded into the original data through an adjacency graph, resulting in flexible projections (Lee and Verleysen 2007). By incorporating the adjacency graph in the process, the solution lies in a relaxed eigenproblem for a sparse matrix, while the PCA, MDS, and their variations solve a rigid eigenproblem for a full matrix. The adjacency graph may be generated using either the $k$ nearest neighbors ($k$NN) or the $\epsilon$-neighborhoods methods. The initial idea of using the adjacency graph to preserve local information was proposed by Roweis and Saul in a DR method called locally linear embedding (LLE) (Roweis and Saul 2000). The LLE is based on the fact that each data point and its neighbors should lie close after DR. To quantify this fact, the LLE constructs a weight matrix with a cost function minimizing the locally linear reconstruction error that occurs when one data point is approximated by its neighbors according to the adjacency graph. Thus, the objective of the projection from the high-dimensional to low-dimensional space is to preserve the reconstruction ability among each data point's neighborhood, which makes the projection locally linear but globally nonlinear. A linear approximation of the LLE method has been later proposed and named neighborhood preserving embedding (NPE) (He et al. 2005). The linear NPE provides a faster and more reliable manner for practical use. With the same intuition to preserve local structures, another graph-based DR called the Laplacian eigenmaps (LE) was proposed by Belkin and Niyogi in 2003 (Belkin and Niyogi 2003). The main advantage of the LE method is the use of the Laplacian matrix on the adjacency graph, which helps to preserve the location information on distance between data points. Also, different from

LLE, the weight matrix in the LE is derived from either a heat kernel or a simple minded 0-1 assignment. In 2003, He and Niyogi proposed the locality preserving projections (LPP) (He and Niyogi 2004). The LPP is a linear approximation of the nonlinear LE. The four mentioned graph-based DR methods (LLE, NPE, LE, and LPP) were initially designed for and mainly used in computer vision context (Teng et al. 2005, Daniel et al. 2011, Wang and Wu 2010, Lewandowski et al. 2010, Hu, Feng and Zhou 2007). In the study of hyperspectral unsupervised DR, the previous graph-based DR researches are limited (Chen and Qian 2007, Ma, Crawford and Tian 2010, Wang and He 2011), as the objective and data-representation are very different between computer vision and hyperspectral images (Han and Goodenough 2005). Some studies combined the graph-based methods with supervised DR methods in hyperspectral applications (Li et al. 2012, Li et al. 2011, Cui et al. 2013), which is beyond the scope of this study.

## 4.1.2    Spectral-spatial dimension reduction methods

Although the local structure in data can be preserved by the graph-based DR methods, the spatial information still is not considered. It is because that these methods are based on linear algebra, which considers a hyperspectral image as a 2D matrix with rows of observations and columns of variables. However, the actual hyperspectral image is a 3D cube with two spatial dimensions along rows and columns indicating the location of the pixel and one spectral dimension indicating spectral bands. By vectorizing the original image along the spectral dimension, the spatial information is lost. Methods based on two different frameworks were proposed to incorporate the spatial information in DR: patch-based and tensor-patch-based analyses. The patch-based methods continue to use the framework from the graph-based methods, but they modified the calculation of adjacency graphs and weight matrices. Instead of only comparing two specific pixels, two patches covering the spatial neighborhood of the two pixels were considered. In this way, the resulting adjacency graphs and weight matrices naturally embrace the spatial information in the process. This patch-based framework was adopted in supervised DR (Zhou et al. 2015, Zhao and Du 2016). However, the patch-based unsupervised DR study has been limited in remote sensing community (Mohan, Sapiro and Bosch 2007).

On the other hand, recently the multilinear algebra is widely being studied. The trend of Datum-as-Is representation sheds light on improved DR methods that preserve the spatial information. The term of tensor means a multidimensional array and the number of dimensions in a tensor can be called as mode or order. A hyperspectral image is a third-order tensor with two spatial modes and one spectral mode. Tensor-based analyses often cover topics of data compression and denoise (Karami, Yazdi and Mercier 2012, Li et al. 2015), but they are seldom of DR. The idea of tensor-based DR appeared around 2010. Several tensor-based DR methods for various computer vision problems have been proposed, including multilinear PCA (Lu, Plataniotis and Venetsanopoulos 2008) and tensor LPP (TLPP) and tensor NPE (TNPE) (He, Cai and Niyogi 2006, Dai and Yeung 2006), concurrent subspaces analysis (Xu et al. 2008), and tensor canonical correlation analysis (Luo et al. 2015). These methods usually work with a group of tensors (e.g. face dataset), while in the case of hyperspectral image DR we only have one tensor. Later, inspired by the patch alignment framework (Zhang et al. 2009), a new tensor-representation has been proposed for hyperspectral images: tensor patch (Zhang et al. 2013, Velasco-Forero and Angulo 2013, Du et al. 2017). Under this tensor-patch-based framework, the input hyperspectral image needs to be first divided into local patches. In this way, the hyperspectral image becomes a 4D dataset, composed by a group of 3D tensors. In (Zhang et al. 2013, Velasco-Forero and Angulo 2013), the tensor patch has been first applied to supervised DR solutions (discriminant locality alignment and linear discriminant analysis). However, the tensor-patch-based unsupervised DR study was limited in remote sensing community. Deng, et al. have adopted the patch alignment framework in the TLPP, which is the first successful unsupervised tensor-patch-based DR method in the field of hyperspectral DR. However, in the paper (Deng et al. 2018), Deng, et al., only provided the experiments on the TLPP, while the solution can be further extended to TNPE.

Regarding the adjacency graph and weight matrix, three methods have been compared (Deng et al. 2018). They are the Euclidean distance, the image patch distance (IPD) (Pu et al. 2014), and the log-Euclidean distance based on the region covariance matrix (RCM) (Tuzel, Porikli and Meer 2006). It was stated that hyperspectral imagery is often

contaminated by noise and the adjacency graph derived directly from the Euclidean distance between original variables may not be accurate. The RCM has been originally proposed in computer vision area to characterize a region of interest as the covariance matrix along the feature dimensions (Tuzel et al. 2006). In the case of the patches from the hyperspectral image, the RCM replaced the original 3D patches with a $d \times d$ 2D matrix, where $d$ is the band number. In the results, the use of RCM outperforms IPD and benefits the preserved information in the DR process. However, the RCM does not consider the fact that in a patch different parts are different effectual in characterizing and recognizing the central pixel. An improved method (weighted RCM (WRCM)) has been proposed and it incorporates a weight matrix in the RCM calculation (Qin et al. 2012). This research adopted the WRCM in constructing the adjacency map/weight matrix, in the hope to improve the DR results.

From the perspective of preserving spatial information, the patch-based framework considers spatial information in the step of constructing an adjacency graph/weight matrix, and the tensor-patch-based framework considers spatial information in both steps of constructing an adjacency graph/weight matrix and solving eigenproblems. Thus, the tensor-patch-based method is a more advanced solution. In regard to space and time complexity, the patch-based and tensor-patch-based methods have the same process of generating adjacency graphs and weight matrices through a group of 3D patches. However, in the next step to create new projections from adjacency graphs and weight matrices, the two frameworks have different space and time complexities. Briefly speaking, the space and time complexities of patch-based methods are separately second-power or third-power correlated with the number of dimensions, and the space and time complexities of tensor-patch-based methods are separately first-power and second-power correlated with the number of samples. For computer vision applications, the dimensions of data are usually much larger than the number of data samples. The tensor-patch-based method may have less intense computation complexity. On the contrary, for hyperspectral images, the dimension of data is only three, when the number of data samples can easily be over tens of thousands. Thus, the tensor-patch-based method faces much larger computation complexity pressure. The question remains about which patch-

based or tensor-patch-based DR solutions for hyperspectral image are better for preserving information and computation complexity.

## 4.2  Objective

The main objective of Chapter 4 is to upgrade the graph-based DR method to its tensor version and further improve the tensor-based method by the use of a new method for the intermediate results of adjacency graph/weight matrix, in the hope to improve the land cover mapping results. Corresponding to the objective, the following research questions are the focus of Chapter 4:

(1) In the two cases of the simulated EnMAP and Hydice urban images, how will the tensor-based DR methods benefit the land cover mapping, compared to the matrix-based DR methods?

(2) What are the differences between the patch-based or tensor-patch-based DR methods? Which one is better?

(3) How does the window size affect the results of the patch-based and tensor-patch-based DR methods?

(4) Does the WRCM derived adjacency map/weight matrix improve the final DR results in land cover mapping?

(5) How can the research results help the current land cover mapping study?

## 4.3  Method

This research tested 16 unsupervised DR methods: LPP, LPP with IPD, LPP with RCM, LPP with WRCM, NPE, NPE with IPD, NPE with RCM, NPE with WRCM, TLPP, TLPP with IPD, TLPP with RCM, TLPP with WRCM, TNPE, TNPE with IPD, TNPE with RCM, and TNPE with WRCM. They root in a same origin and can be considered as evolved variations to the classical methods. This section will provide a detailed explanation of each of the methods and their relations.

## 4.3.1　Dimension reduction methods

Two groups of DR methods will be studied in this research: patch-based (LPP and NPE related) and tensor-patch-based (TLPP and TNPE related) methods. They both account for the spectral and spatial information in the data using different strategies.

### 4.3.1.1　LPP and NPE

The LPP (locality preserving projections) and NPE (neighborhood preserving embedding) methods are two of the early attempts in adjacency-graph-based methods. They separately simplified the nonlinear LE and LLE methods and embrace great simplicity and flexibility. They have the same processing manner of three steps, but with different objective functions. The three steps are as follows:

(1) Constructing the adjacency graph $G$. For a 1D dataset $X$ with $n$ data points, each data point becomes a node in the adjacency graph. An edge will be built between two nodes if they are "neighbors". The most common way to decide the neighborhood of one node is the $k$ nearest neighbors (KNN) method. The KNN method connects two nodes if they are among the $k$ nearest neighbors of each other using Euclidean (Zhong et al. 2015, Gao et al. 2015) distance.

(2) Generating the weight matrix $W$. The weight matrix is a sparse matrix with a size of $n \times n$. In LPP, the values in the weight matrix are commonly calculated by a heat kernel and are 0 if two nodes are not connected (He and Niyogi 2004):

$$w_{ij} = e^{-\frac{\|x_i - x_j\|^2}{t}}. \tag{4-1}$$

where $\|\|$ is the Euclidean norm. In NPE, the values in the weight matrix are calculated by minimizing the function below (He et al. 2005):

$$\min \sum_i \| x_i - \sum_j w_{ij} x_j \|^2, \tag{4-2}$$

with constraints $\sum_j w_{ij} = 1, j = 1,2, \dots, n$.

(3) Solving the eigenfunction. The objective of the LPP is to ensure that if two data points are close to each other in the original space, they should stay close in the projected feature space. It can be realized through the minimization problem below:

$$\min \sum_{ij}(y_i - y_j)^2 w_{ij}, \tag{4-3}$$

where $y_i$ and $y_j$ are data points in the projected feature space. He and Niyogi justified that the above minimization problem is equivalent to solve the eigenfunction below, whose derivation can be found in Appendix C (He and Niyogi 2004):

$$XLX^T \boldsymbol{a} = \lambda XDX^T \boldsymbol{a}, \tag{4-4}$$

where the Laplacian matrix $L = D - W$; and $D_{ii} = \sum_j W_{ji}$ is a diagonal matrix with column sums of W.

In the case of NPE, the objective function is to keep the reconstruction error among each neighborhood small:

$$\min \sum_{ij}(y_i - \sum_j w_{ij} y_j)^2. \tag{4-5}$$

This minimization problem is also simplified as an eigenfunction:

$$XMX^T \boldsymbol{a} = \lambda XX^T \boldsymbol{a}, \tag{4-6}$$

where $M = (I - W)^T(I - W)$ and $I = diag(1, \dots, 1)$.

The derived eigenvector $\boldsymbol{a}$ is the resulting projection. In the case of hyperspectral images, the original image is reordered into a 1D data along the spectral dimension with $n$ pixels. The flowchart of the LPP and NPE methods is shown in Figure 4-1, where the letters in red show the calculation method of the adjacency map and weight matrix.

**Figure 4-1: Flowchart of the LPP and NPE methods.**

## 4.3.1.2    Tensor LPP and Tensor NPE

The LPP and NPE methods are able to explore the spectral features of the image in both global and local ranges. But, they failed to consider any of the spatial information, as the two spatial modes in the hyperspectral image are integrated as one observation mode. Due to the increasingly popular multilinear algebra, the tensor representation of data provides a natural way to preserve the spatial information. With proper modification, the NPE and LPP methods have been upgraded into their tensor versions, separately called tensor NPE (TNPE) and tensor LPP (TLPP) (Dai and Yeung 2006, He et al. 2006). The TNPE and TLPP methods keep the original dimensions of the data and apply individual NPE and LPP on each dimension sequentially.  The TLPP and TNPE have fulfilled this goal. One basic tensor terminology and one tensor operations that will be used in the TLPP and TNPE are explained below.

Definition1 Given a $n$-mode tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n}$, the $f$-mode unfolding of tensor $\mathcal{A}$ is denoted as $\mathcal{A}^{(f)}$. It flattens the tensor $\mathcal{A}$ into a matrix $\mathcal{A}^{(f)} \in \mathbb{R}^{I_f \times I_1 \cdots I_{f-1} I_{f+1} \cdots I_n}$. The columns of $\mathcal{A}^{(f)}$ is obtained by fixing all but one mode.

Definition 2 A $f$-mode product between a $n$-mode tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n}$ and a matrix $U \in \mathbb{R}^{J \times I_f}$ gives a tensor $\mathcal{A} \times_f U \in \mathcal{R}^{I_1 \times \cdots I_{f-1} \times J \times I_{f+1} \cdots \times I_n}$. The element presentation of the $f$-mode product is

$$\mathcal{A} \times_f U_{i_1 \cdots I_{f-1} j i_{f+1} \cdots i_n} = \sum_{i_f=1}^{I_f} \mathcal{A}_{i_1 i_2 \cdots i_n} U_{j i_f}. \tag{4-7}$$

Using the Definition 1, the $f$-mode unfolding of the $f$-mode product results can be expressed as below:

$$(\mathcal{A} \times_f U)^{(f)} = U \mathcal{A}^{(f)}. \tag{4-8}$$

The TLPP and TNPE methods are also realized through three similar steps like the LPP and NPE methods, but with some modifications to account for the tensor representation of the data. The application of the TLPP and TNPE in hyperspectral DR is limited. It is probably because of the inexplicit tensor representation of hyperspectral data. Here, we will explain the TLPP and TNPE with a case of a third-order tensor representing a hyperspectral image. The given 3D hyperspectral image $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$ is spatially segmented into multiple 3D patches with the same window size $m$ $\mathcal{X}_q \in \mathbb{R}^{m \times m \times K}$, where $I$ and $J$ are the row and column numbers representing the spatial location and $K$ is the spectral band number. This process creates a new 4D dataset from the original data. Then a certain number $n$ of training patches are selected as input to the TLPP/TNPE algorithms with the following three steps:

(1) Constructing the adjacency graph $G$. Each of the $n$ training cubes from the hyperspectral image becomes a node in the adjacency graph. An edge will be built between two nodes if they are "neighbors". The most common way to decide the neighborhood of one node is the $k$ nearest neighbors (KNN) method.

(2) Generating the weight matrix $W$. The TLPP and TNPE uses the same methods correspondingly as the above LPP and NPE methods.

(3) Solving the eigenfunction. Like the LPP, the objective function of the TLPP method also aims to ensure that if two data points are close to each other in the original space, they should stay close in the projected feature space. It can be realized through the minimization problem below (Dai and Yeung 2006):

$$\min \sum_{ij} (\mathcal{Y}_i - \mathcal{Y}_j)^2 W_{ij}, \tag{4-9}$$

where $\mathcal{Y}_i$ and $\mathcal{Y}_j$ are hyperspectral training cubes in the projected feature space. This minimization problem can be solved by the eigenfunction below:

$$\left( \sum_{ij} W_{ij} (\mathcal{Y}_i^{(f)} - \mathcal{Y}_j^{(f)}) (\mathcal{Y}_i^{(f)} - \mathcal{Y}_j^{(f)})^T \right) \mathbf{U}_f =$$
$$\lambda \left( \sum_i \mathcal{Y}_i^{(f)} \mathcal{Y}_i^{(f)T} D_{ii} \right) \mathbf{U}_f, f = 1, 2, 3, \tag{4-10}$$

where $D_{ii} = \sum_j W_{ji}$.

The TNPE method has the same objective as to keep the reconstruction error among each neighborhood small:

$$\min \sum_i \| \mathcal{Y}_i - \sum_j W_{ij} \mathcal{Y}_j \|, \tag{4-11}$$

The above minimization problem can be solved by the eigenfunction below:

$$\left( \sum_i (\mathcal{Y}_i^{(f)} - \sum_j W_{ij} \mathcal{Y}_j^{(f)}) (\mathcal{Y}_i^{(f)} - \sum_j W_{ij} \mathcal{Y}_j^{(f)})^T \right) \mathbf{U}_f =$$
$$\lambda \left( \sum_i \mathcal{Y}_i^{(f)} \mathcal{Y}_i^{(f)T} \right) \mathbf{U}_f, f = 1, 2, 3, \tag{4-12}$$

For both TLPP and TNPE, the eigenfunction is solved along each of the three dimensions to achieve the projections $\mathbf{U_1}$, $\mathbf{U_2}$, and $\mathbf{U_3}$. The flowchart of the TLPP and TNPE methods is shown in Figure 4-2, where the letters in red show the calculation method of the adjacency map and weight matrix.
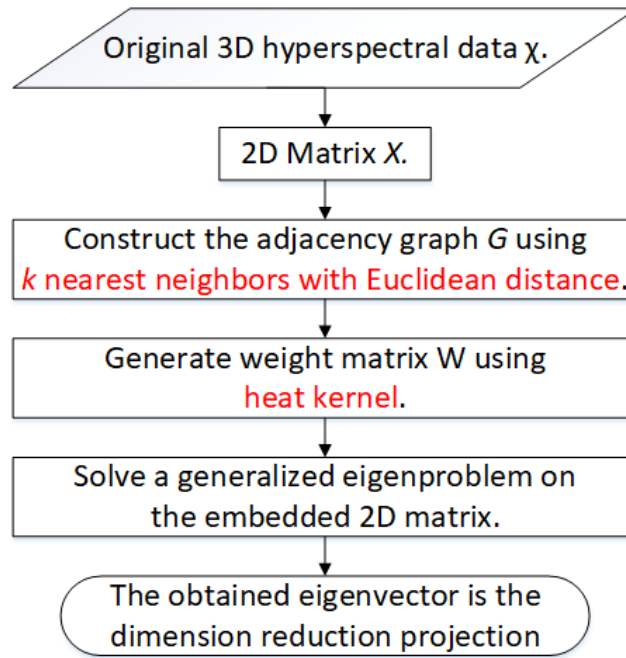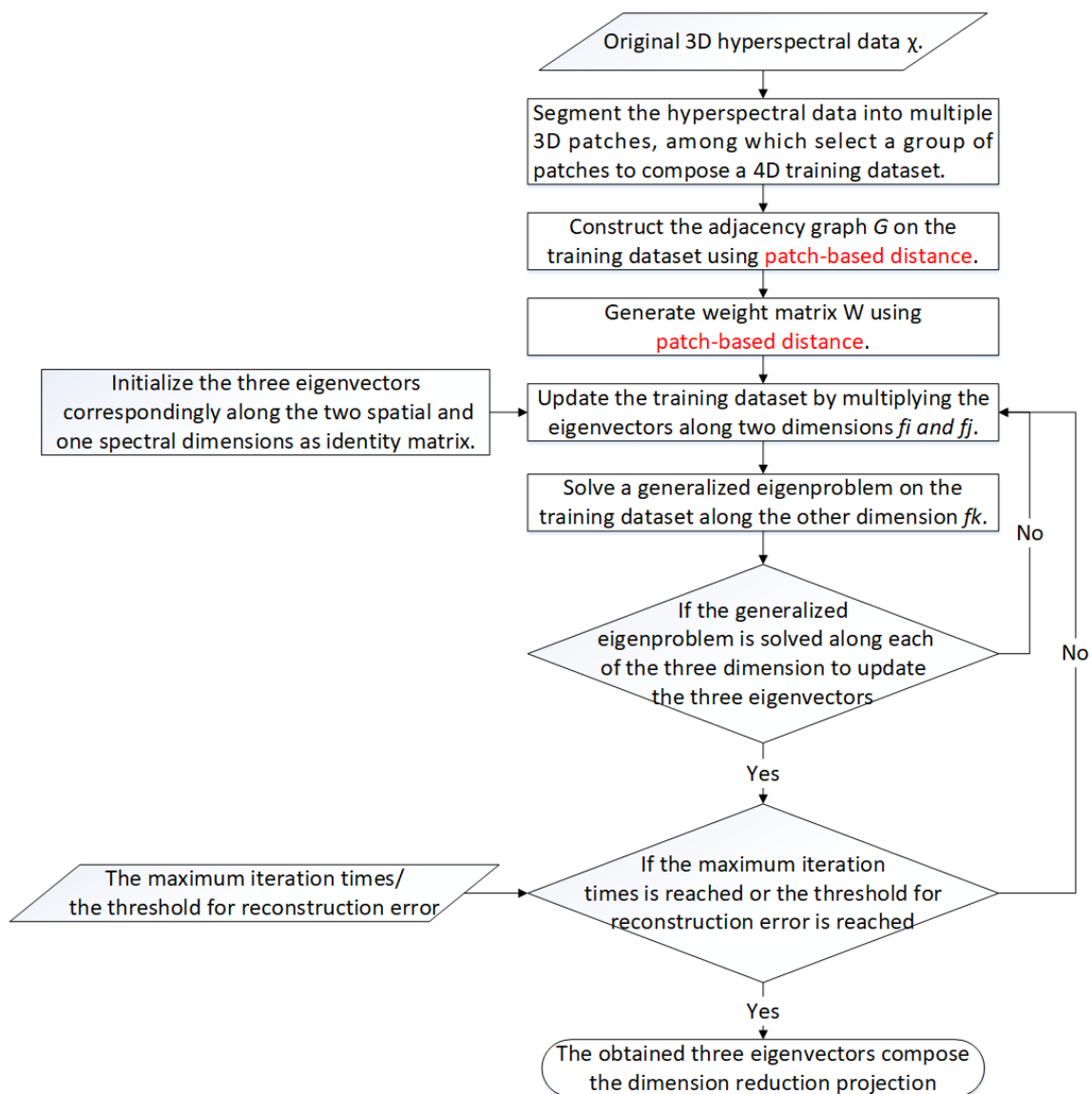
**Figure 4-2: Flowchart of the TLPP and TNPE methods.**

## 4.3.2     Patch-based adjacency graphs and weight matrices

From the above explanations on LPP, NPE, TLPP, and TNPE methods, one can easily understand the important role played by the adjacency graph and weight matrix. The effectiveness of the DR heavily depends on the information captured in the adjacency graph and weight matrix. In the traditional graph-based methods, the original data are used to calculate the adjacency graph and weight matrix. Three different methods

116

considering both spectral and spatial information in the data were used in this research to produce adjacency graphs and weight matrices: IPD, RCM, and WRCM. The IPD has been exclusively proposed for hyperspectral similarity measure. The RCM was newly introduced to hyperspectral DR from computer vision. The WRCM is a newer method developed from RCM.

## 4.3.2.1 Moving window

Before introducing the three spatial-spectral methods, we first consider the moving window cases for three different pixel locations: at the corner, on the edge and in the image. For a given pixel, a group of $w \times w$ pixels should be decided as the spatial neighbors, where $w$ is the window size. If the pixel is located in the image, the moving window can naturally cover its spatial neighbors. If the pixel lies on the edges or at the corner of the image, we used a reflection transformation to fill the non-existing spatial neighbors. Examples of the three pixel locations are shown in Figure 4-3.



| $X_{22}$ | $X_{21}$ | $X_{22}$ |
|---|---|---|
| $X_{12}$ | $X_{11}$ | $X_{12}$ |
| $X_{22}$ | $X_{21}$ | $X_{22}$ |

(a)

| $X_{2(j-1)}$ | $X_{2j}$ | $X_{2(j+1)}$ |
|---|---|---|
| $X_{1(j-1)}$ | $X_{1j}$ | $X_{1(j+1)}$ |
| $X_{2(j-1)}$ | $X_{2j}$ | $X_{2(j+1)}$ |

(b)

| $X_{(i-1)(j-1)}$ | $X_{(i-1)j}$ | $X_{(i-1)(j+1)}$ |
|---|---|---|
| $X_{i(j-1)}$ | $X_{ij}$ | $X_{i(j+1)}$ |
| $X_{(i+1)(j-1)}$ | $X_{(i+1)j}$ | $X_{(i+1)(j+1)}$ |

(c)
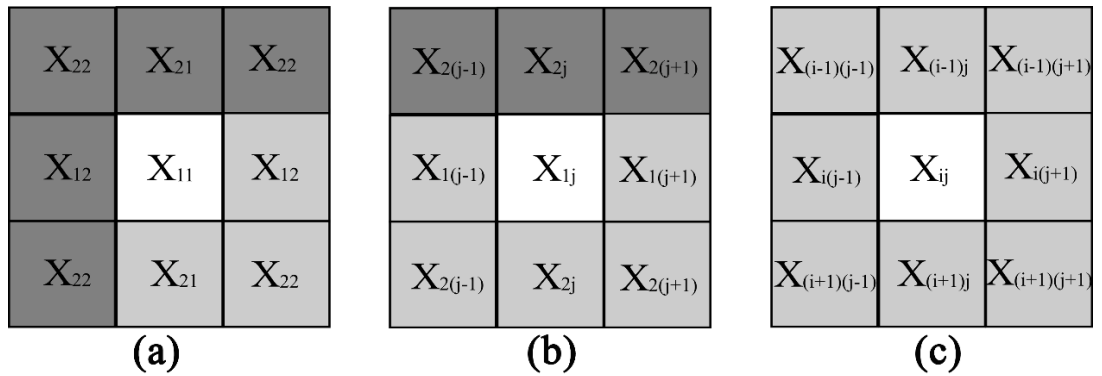
**Figure 4-3: Spatial neighbors of three cases**
**(a) At the corner, (b) On the edge, and (c) In the image.**

## 4.3.2.2 Image patch distance

The IPD is specially proposed to be used in adjacency-graph based methods like LPP and NPE. This method takes advantage of both spatial and spectral information in the data. For any two pixels in the hyperspectral image, the IPD calculate their similarity based on the small neighborhood (spatial window) of the two pixels. Given two pixels $x_i$ and $x_j$

from a hyperspectral image, a spatial window of size $w$ is used to find the neighborhoods for the two pixels $x_i$ and $x_j$. The neighborhoods are correspondingly denoted as $\Omega(x_i) = \{a_1, a_2, \dots, a_{w^2}\}$ and $\Omega(x_j) = \{b_1, b_2, \dots, b_{w^2}\}$. The IPD is defined as

$$d_{IPD}(x_i, x_j) = \sum_{l=1}^{w^2} d_u(a_l, b_l),$$

$$d_u(a_l, b_l) = \max(\min_{b \in \Omega(x_j)} d(a_l, b), \min_{b \in \Omega(x_i)} d(b_l, a)),$$

(4-13)

where $d(a, b)$ is a spectral similarity function comparing a to b. Using this equation, the similarity between two pixels is measured with their surrounding neighbors. Thus, the IPD incorporates both the spatial and spectral information. Figure 4-4 shows the calculation of IPD with a $3 \times 3$ spatial window size.
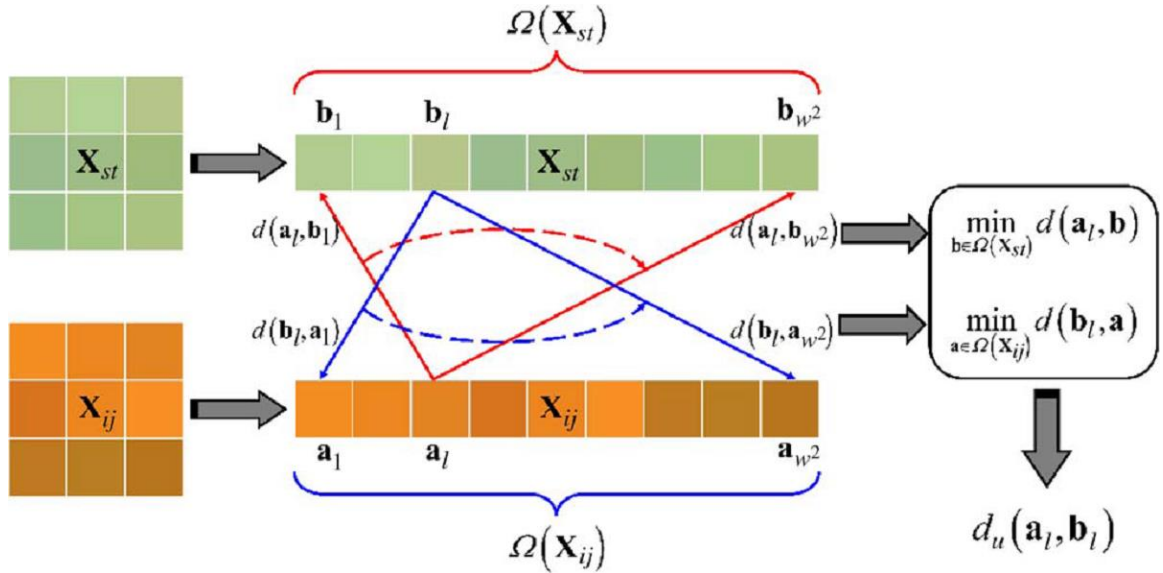


**Figure 4-4: Flowchart to calculate the IPD with a $3 \times 3$ spatial window (Pu et al. 2014).**

### 4.3.2.3    Region covariance matrix

The RCM is an advantageous region descriptor proposed by Tuzel, et al. This method describes a given region of interest with the covariance matrix of the feature points.

Given a region of interest $R \in \mathbb{R}^{m \times m \times K}$, where $m$ is the window size, from the hyperspectral image $\mathcal{X} \in \mathbb{R}^{I \times J \times K}$, it can be represented by a $K \times K$ covariance matrix:

$$C_R = \frac{1}{n}\sum_{i=1}^{n}(r_i - \mu)(r_i - \mu)^T, \qquad (4\text{-}14)$$

where $n = m \times m$ is the number of pixels in the regions of interest; $r_i$ is the $K$-dimensional feature point in R; and $\mu = \frac{1}{n}\sum_{i=1}^{n}r_i$. The resulting RCM is a symmetric matrix along the diagonal. The RCM calculates the correlation among features in a region of interest and has no information regarding the location and number of pixels. It has a few properties: if two features ($a$ and $b$) tend to increase together, then $C_R(a,b) > 0$; if feature $a$ tends to decrease when feature $b$ increases, then $C_R(a,b) < 0$; and if two features ($a$ and $b$) are independent, then $C_R(a,b) = 0$. In computer vision, the commonly used image features include pixel locations *(x, y)*, color (RGB) values, intensity, and the norm of the first and second order derivatives of the intensities with respect to *x* and *y*. An example of two orthophotos separately covering one grass area and one residential area is shown in Figure 4-5.

$$F(x,y) = [x, y, I_R(x,y), I_G(x,y), I_B(x,y), \partial_x I(x,y), \partial_y I(x,y)]$$
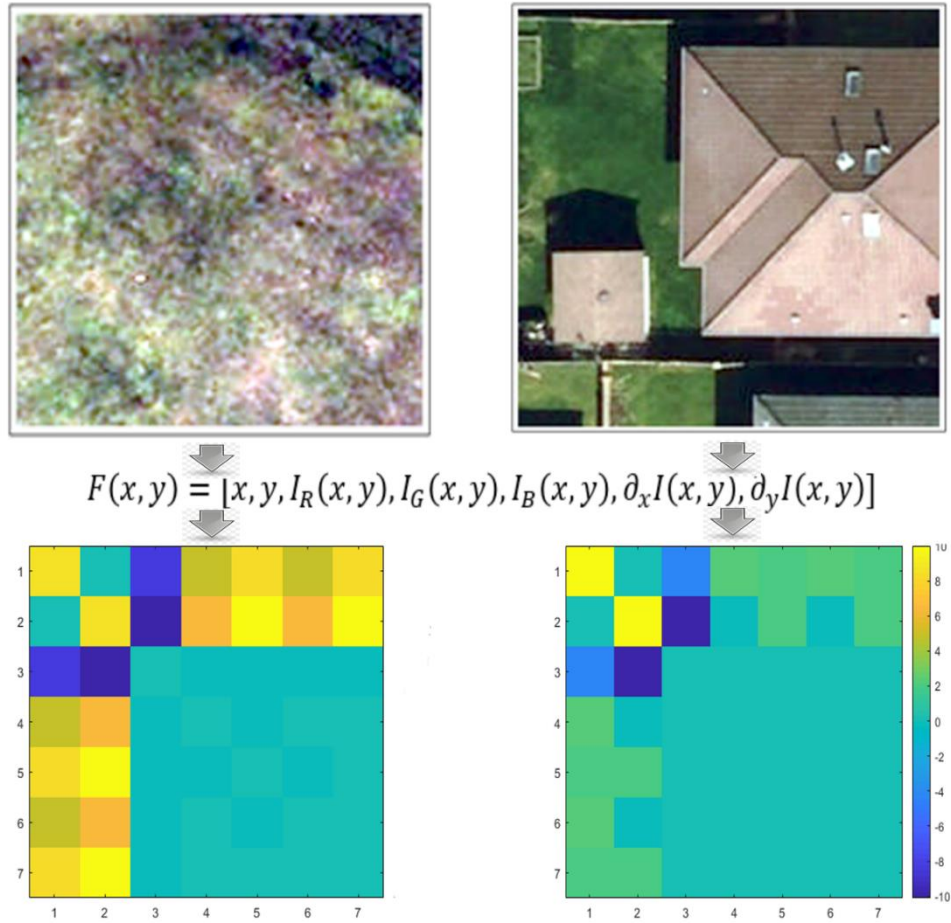
**Figure 4-5: Example results from RCM.**

Since covariance features lie on a Rimannian manifold, a log-Euclidean distance is chosen to compute the distance between two RCM:

$$D_{LE}(C_i, C_j) = \| \log(C_i) - \log(C_j) \|, \qquad (4\text{-}15)$$

where $log$ represents the matrix logarithm.

## 4.3.2.4    Weighted region covariance matrix

The WRCM was proposed as an improved version of RCM. Reformulates Equation (4-14) from RCM as (Qin et al. 2012):

$$C_W = \frac{1}{2} \frac{1}{n \times n} \sum_{i=1}^{n} \sum_{j=1}^{n} (r_i - r_j)(r_i - r_j)^T, \tag{4-16}$$

by replacing $\mu$ as:

$$\mu = \frac{1}{n} \sum_{i=1}^{n} r_i, \tag{4-17}$$

From Equation (4-16), we can see that each pair of pixels in the current patch contributes equally to the final results, which fails to reflect the spatial distribution in the patch. The WRCM method introduces a weight term in the calculation:

$$C_W = \frac{1}{2} \frac{1}{n \times n} \sum_{i=1}^{n} \sum_{j=1}^{n} (r_i - r_j)(r_i - r_j)^T w_{ij}, \tag{4-18}$$

where $w_{ij}$ is calculated as Equation (4-1). Equation (4-18) can be further deduced as below:

$$C_W = XLX^T, \tag{4-19}$$

where the Laplacian matrix $L = D - W$; $D_{ii} = \sum_j W_{ji}$ is a diagonal matrix with column sums of W; and $X$ is the original patch. The measure of distances between weighted covariance matrices are adopted from the original RCM method (Tuzel et al. 2006) and is based on an eigenproblem:

$$D(C_i, C_j) = \sqrt{\sum_{i=1}^{n} ln^2 \lambda_i(C_1, C_2)}, \tag{4-20}$$

where $\lambda_i(C_1, C_2)$ is the generalized eigenvalues of $C_1$ and $C_2$, computed from

$$\lambda_i C_1 x_i - C_2 x_i = 0, i = 1, \dots, d \tag{4-21}$$

where $x_i$ are the generalized eigenvectors. The resulting WRCM is a symmetric matric along the diagonal. Similar to the RCM, the WRCM calculates the correlation among features in a region of interest and has no information regarding the location and number of pixels. Based on the RCM, the WRCM reflects the weighted feature correlation, when

those feature correlations of pixels similar to the center pixel exert more effect on the results than those non-similar pixels. An example of two orthophotos separately covering one grass area and one residential area is shown in Figure 4-6.



$$F(x,y) = [x, y, I_R(x,y), I_G(x,y), I_B(x,y), \partial_x I(x,y), \partial_y I(x,y)]$$
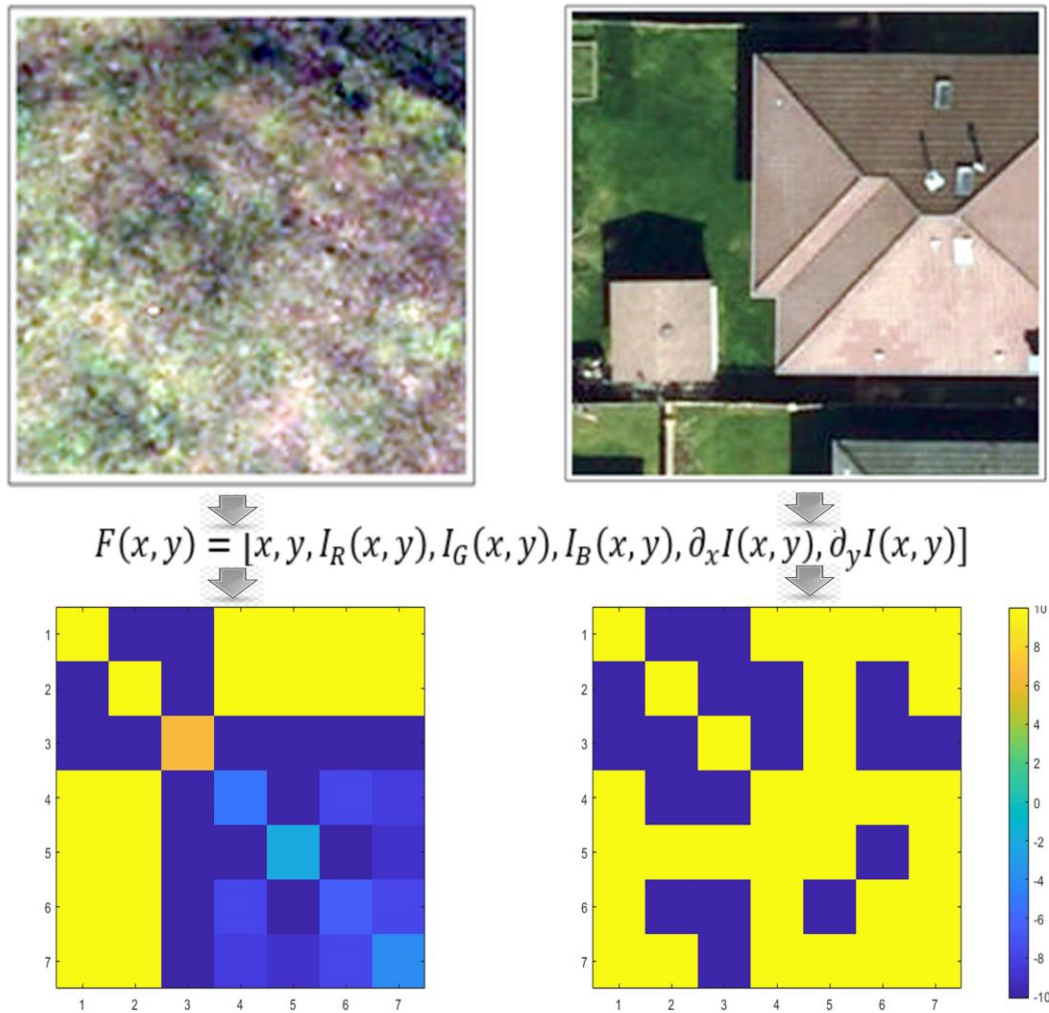
**Figure 4-6: Example results from WRCM.**

## 4.3.3    Computation complexity

The construction of adjacency graphs and weight matrices is one major part in the DR methods used in this research. From the above descriptions, the computational complexity and memory consumption of this part is very considerable, especially for hyperspectral image DR. Specifically, for calculating the adjacency graphs and weight

matrices using the original image without patch-based modifications, the computation complexities are separately $O(dn^2)$ and $O(dnk^3)$ (Chen and Qian 2007), where d is the input dimensionality, k the number of nearest neighbours, $n$ is the number of data points. For hyperspectral images, the value of d can easily reach 100 to 200, and the value of $n$ in testing data is usually $512 \times 512$ or $1024 \times 1024$. While in real-life cases, it can be even larger. Furthermore, in order to consider the spatial information for each pixel, patch-based adjacency graphs and weight matrices are calculated with much higher complexity. The distance calculation between each pair of data points is a big burden in this process. The heavy computation may be another reason of the limited applications of the graph-based methods in hyperspectral studies. In TLPP and TNPE, this problem is solved by selecting a group of training samples to derive the resulting projections, instead of using the entire dataset. This research adopts this solution for both LPP/ NPE- and TLPP/TNPE- based methods as 20% of data was randomly selected as training sample for DR. In order to further accelerate the process, this research adopted a parallel computation manner provided by the parallel computing toolbox from Matlab. The parallel computation allows multiple sessions to simultaneously calculate blocks of distances. In the case of the RCM and WRCM methods, the covariance matrices calculation for hundreds of bands consumes a large amount of time. The traditional RCM method used in computer vision often processes intensity, gradient, and Laplacian magnitude of the original image (Tuzel et al. 2006). Later, improved RCM methods tend to use Gabor features of the image to calculate covariance, which further enhance the discrimination ability (Pang, Yuan and Li 2008, Qin et al. 2012). In order to efficiently obtain the covariance matrix in RCM and WRCM methods, this study used the first three principal components of the original image to create RCM and WRCM.

## 4.3.4    Evaluation of the dimension reduction method

In order to evaluate the DR (dimension reduction) results, classification was performed on the series of dimension-reduced images. The support vector machine (SVM) classifier was used in this study. As benchmark, we also applied the SVM classifier on the original image and the PCA dimension-reduced result. We used two testing hyperspectral images

with different spatial resolution. The two images are separately located in urban and suburban environments, focusing on targets of different scales. The differences between the two images allow us to compare the impact of different images on the choice of window size in the calculation of the adjacency graphs and weight matrix. In the results, we generated the classification confusion matrices, providing the overall accuracy, producer's accuracy, and user's accuracy.

## 4.4 Experiments and results

In this section, we will first illustrate the overall design of our experiments and the hyperspectral images used in the analysis. The results of the DR and classification are shown.

### 4.4.1    Studied hyperspectral images and setup

This study used two hyperspectral images. The first hyperspectral image depicts the urban area in City of Surrey, BC, Canada (Figure 4-7). It was obtained by the airborne CASI-1500 sensor during April 2013. The image contains 72 spectral bands from visible to near-infrared portion (0.36 μm to 1.05 μm) with a 9.6 nm band interval. The spatial resolution of the image is 1m. We selected a small part from the CASI hyperspectral data of size $150 \times 150$. The image is located in a mixed commercial and residential area, containing large impervious surface area as well as green spaces. The data was pre-processed prior to use in this study. The three major pre-processing steps include the following: 1) radiometric and spectral calibration converting raw brightness values into spectral radiance values; 2) georeference by an inertial measurement unit (IMU)[5] and a GPS system; and 3) atmospheric adjustment to remove effects of atmospheric absorption and convert radiance to reflectance data. The second image is the widely used test data: Indian Pines (Figure 4-8). The Indian Pines dataset is collected by the AVIRIS sensor in 1992 over an agricultural area in Northern Indiana, IN, USA. The spectral bands of the Indian Pines data span from 400-2500 nm. After deleting the 20 bands that are affected by water absorption, 200 bands were used in this study. The spatial resolution is 20m.

---

[5] The IMU is a device that records aircraft motion and attitude.

The Indian Pines image has a size of $145 \times 145$ and contains 10249 samples in 16 different classes.
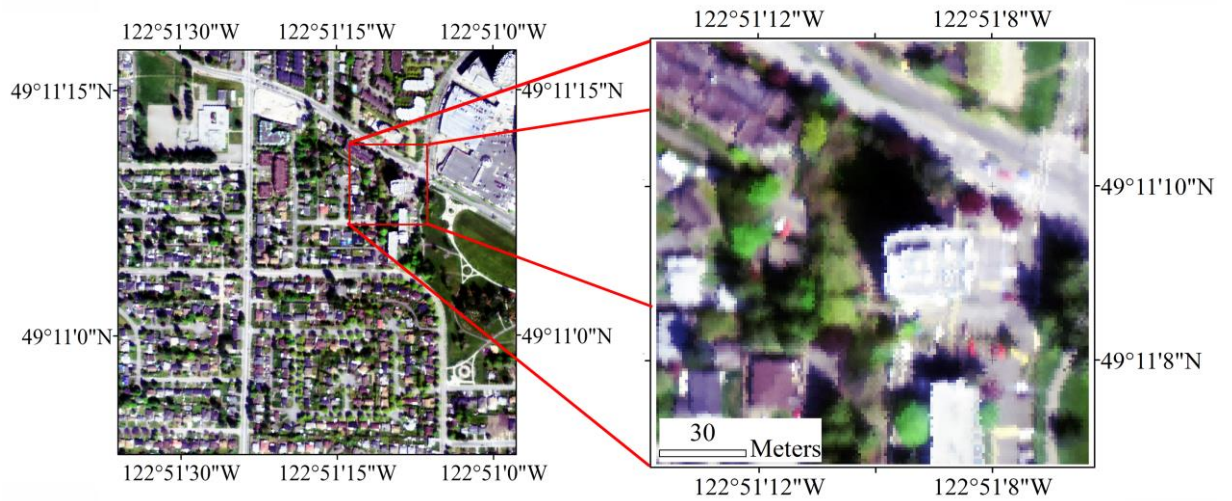


**Figure 4-7: Surrey, BC, CASI hyperspectral image in RGB.**



**Figure 4-8: Indian Pines, AVIRIS hyperspectral image in RGB.**

To fulfill the objectives of this research, we designed our experiment as follows. Among the 16 tested DR methods, half of them are patch-based methods (LPP, LPP with IPD, LPP with RCM, LPP with WRCM, NPE, NPE with IPD, NPE with RCM, NPE with WRCM), and half of them are tensor-patch-based methods (TLPP, TLPP with IPD, TLPP with RCM, TLPP with WRCM, TNPE, TNPE with IPD, TNPE with RCM, and TNPE with WRCM). In order to obtain an appropriate local window size and preserved dimensions, several cross-validation experiments were performed respectively. For the 14 methods that apply the original data, IPD, RCM, or WRCM, the window sizes are correspondingly set as $3 \times 3$, $5 \times 5$, $7 \times 7$, $9 \times 9$, $11 \times 11$, and $13 \times 13$. From now on, the three adjacency map/weight matrix methods may be represented by its name followed by a parenthesis with a number, e.g. IPD (3), where the number in a parenthesis indicates the window size. A set of preserved dimensions (2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, and 30) was chosen. The dimension-reduced images were then classified using SVM classifier. For the first studied image, we manually selected training samples, covering six land cover types: trees, grasses, concrete, asphalt, and dark shingles. Then, we used the orthophoto data taken one month earlier than the CASI hyperspectral data in the same location as reference data. The orthophoto data has a 10 cm spatial resolution and contains three bands of RGB. In the case of the second studied image, reference data are available on GIC's website [6] (Computational Intelligence Group). In the suburban scene, 16 land cover types were identified: alfalfa, corn (no till), corn (minimal till), corn, grass-pasture, grass-tree, grass-pasture-mowed, hay-windrowed, oats, soybean (no till), soybean (minimal till), soybean-clean, wheat, woods, building-grass-tree-drive, and stone-steel-tower. We randomly selected training samples from the original Indian Pines, AVIRIS image, and the remaining pixels act as testing samples.

---

[6]http://www.ehu.eus/ccwintco/index.php?title=P%C3%A1gina_principal

## 4.4.2 Dimension reduction results

## 4.4.2.1 Patch-based dimension reduction results

In the case of patch-based DR, eight methods were processed (LPP, LPP with IPD, LPP with RCM, LPP with WRCM, NPE, NPE with IPD, NPE with RCM, NPE with WRCM). Except LPP and NPE, the other six methods were repeatedly applied using six different window sizes to calculate the patch-based adjacency graphs and weight matrices. Thus, in total there are 38 dimension-reduced images for each studied hyperspectral image. Based on the following classification results, for each patch-based method, the first band of the DR result with the window size that has the best overall accuracy, shown below separately for the Surrey, CASI (Figure 4-9) and Indian Pines, AVIRIS (Figure 4-10) images. The first band of the DR result should preserve the most amount of useful information from the original image. One noticeable difference between the LPP-based and NPE-based methods is that the patch-based LPP methods have a fuzzy effect over the image. This fuzzy effect happens mainly among unified areas of the same land cover types, while the boundaries between different land cover types are well kept with less of a fuzzy effect. Compared to RCM- and WRCM-based LPP DR method, the IPD-based LPP method delivers the fuzziest results.



| LPP | LPP+IPD (3) | LPP+RCM (7) | LPP+WRCM (7) |
| NPE | NPE+IPD (9) | NPE+RCM (11) | NPE+WRCM (13) |
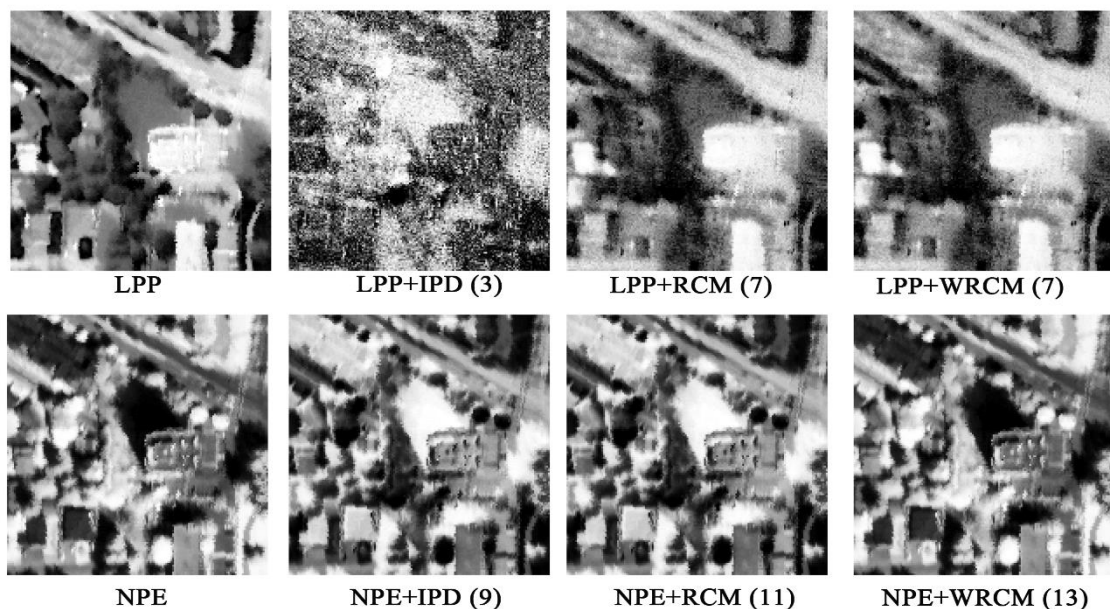
**Figure 4-9: First band of the dimension-reduced images with the window size (in parentheses) that has the best overall accuracy (CASI image).**



LPP          LPP+IPD (3)          LPP+RCM (13)          LPP+WRCM (13)

NPE          NPE+IPD (3)          NPE+RCM (11)          NPE+WRCM (11)
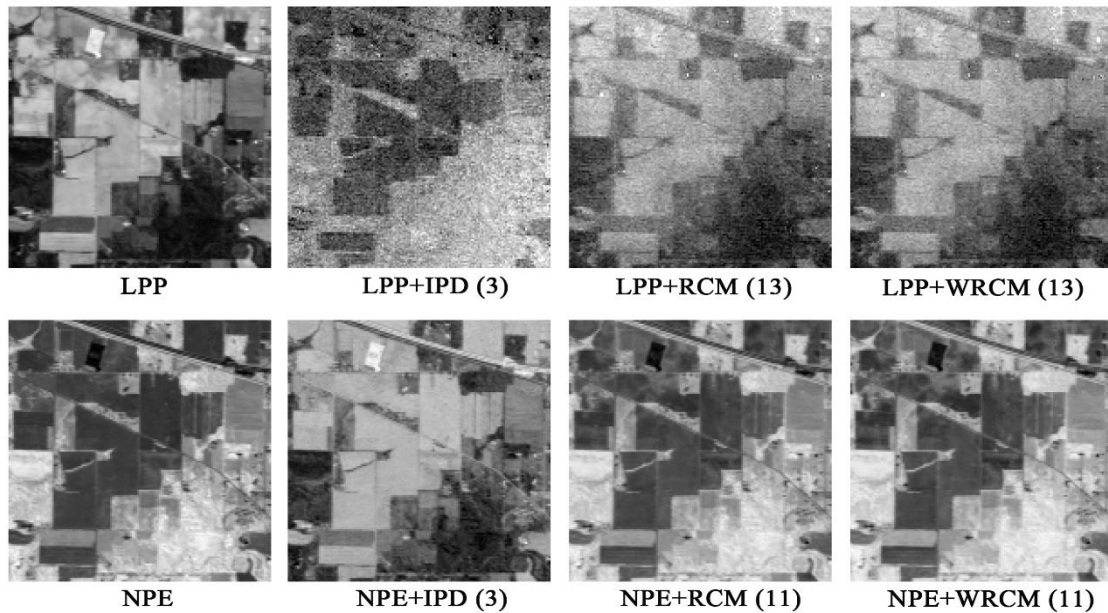
**Figure 4-10: First band of the dimension-reduced images with the window size (in parenthesis) that has the best overall accuracy (AVIRIS image).**

## 4.4.2.2    Tensor-patch-based dimension reduction results

In the case of tensor-patch-based DR, there are also eight methods (TLPP, TLPP with IPD, TLPP with RCM, TLPP with WRCM, TNPE, TNPE with IPD, TNPE with RCM, and TNPE with WRCM). The eight tensor-patch-based DR methods were repeatedly applied using six different window sizes to calculate the patch-based adjacency graphs and weight matrices. Thus, in total there are 48 dimension-reduced images for each studied hyperspectral image. Based on the following classification results, for each tensor-patch-based method, the first band of the DR result with the window size that has the best overall accuracy was shown below separately for the Surrey, CASI (Figure 4-11) and Indian Pines, AVIRIS (Figure 4-12) images. Compared to the patch-based results, the tensor-patch-based DR results are fuzzier. In both of the studied images, the modified TLPP and TNPE methods that use the IPD-, RCM-, and WRCM-produced adjacency

graph/weight matrix generate fuzzier images than the traditional TLPP and TNPE methods. Different from the patch-based methods, the IPD-based TLPP and TNPE resulted in the least fuzzy image compared to the RCM- and WRCM-based TLPP and NPE methods.



TLPP (5)　　　　TLPP+IPD (13)　　　　TLPP+RCM (7)　　　　TLPP+WRCM (7)

TNPE (5)　　　　TNPE+IPD (3)　　　　TNPE+RCM (7)　　　　TNPE+WRCM (9)
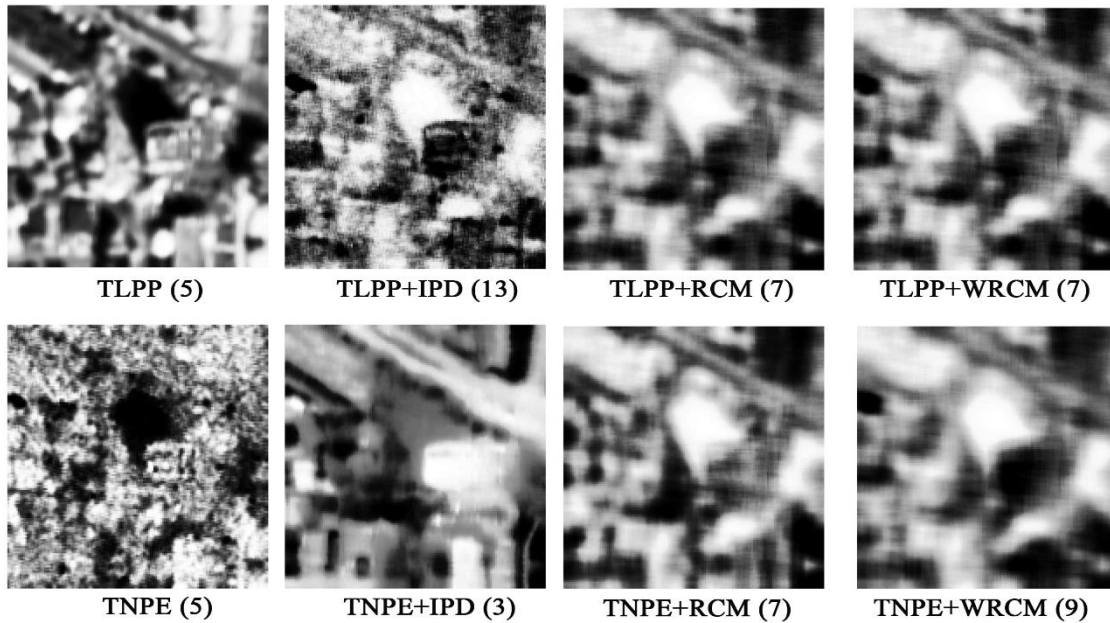
**Figure 4-11: First band of the dimension-reduced images with the window size (in parentheses) that has the best overall accuracy (CASI image).**
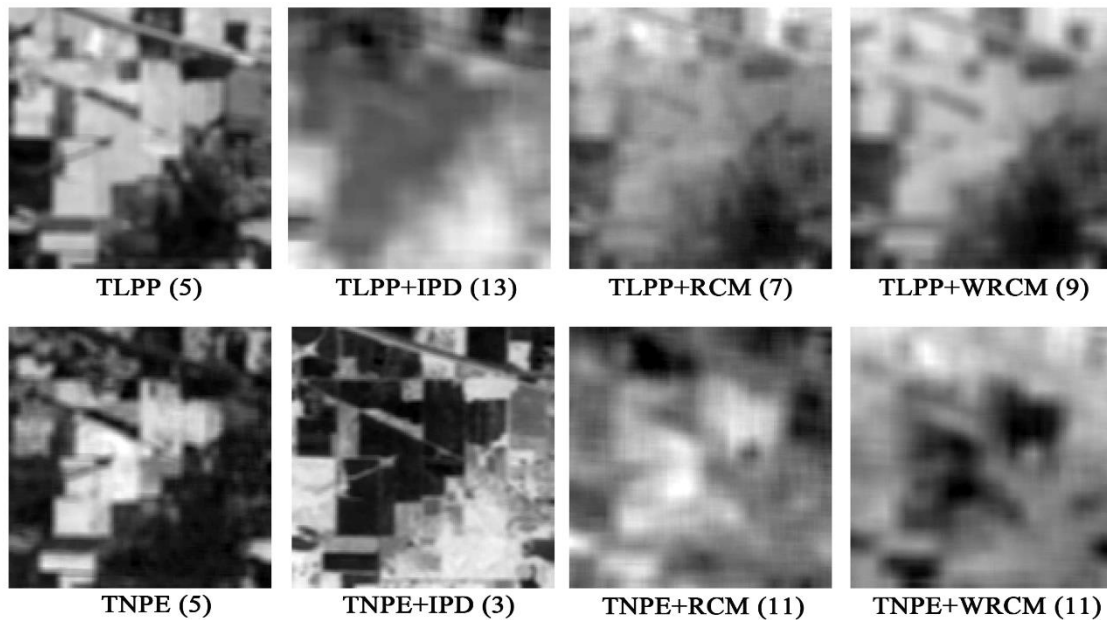
**Figure 4-12: First band of the dimension-reduced images with the window size (in parentheses) that has the best overall accuracy (AVIRIS image).**

## 4.4.3    Classification results

In order to determine the effectiveness of the DR methods, we used the SVM classifier to generate a series of classification maps. A group of 471 training samples and a group of 2610 training samples were separately generated for the Surrey, BC, CASI and Indian Pines, AVIRIS images respectively (Figure 4-11 and 4-12). For comparison, the SVM classification results of the original image and the PCA dimension-reduced results were also generated. In the case of PCA DR, classification was performed with 15 sets of preserved bands (2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, and 30). To evaluate the classification results, one set of 13740 and one set of 7288 testing points were generated separately for the Surrey, BC, CASI and Indian Pines, AVIRIS images (Figure 4-13 and 4-14).
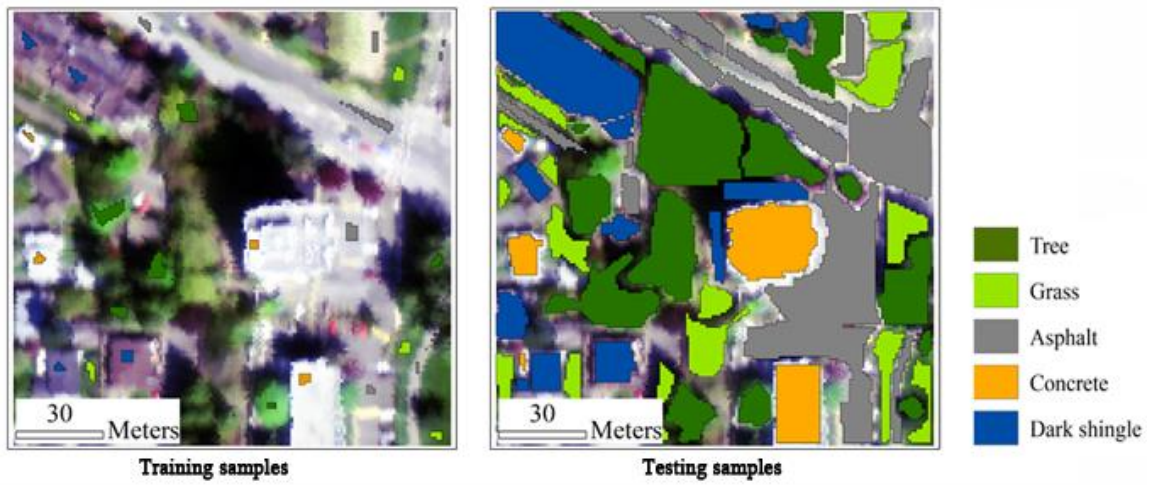
**Figure 4-13: Training and testing samples (CASI image).**
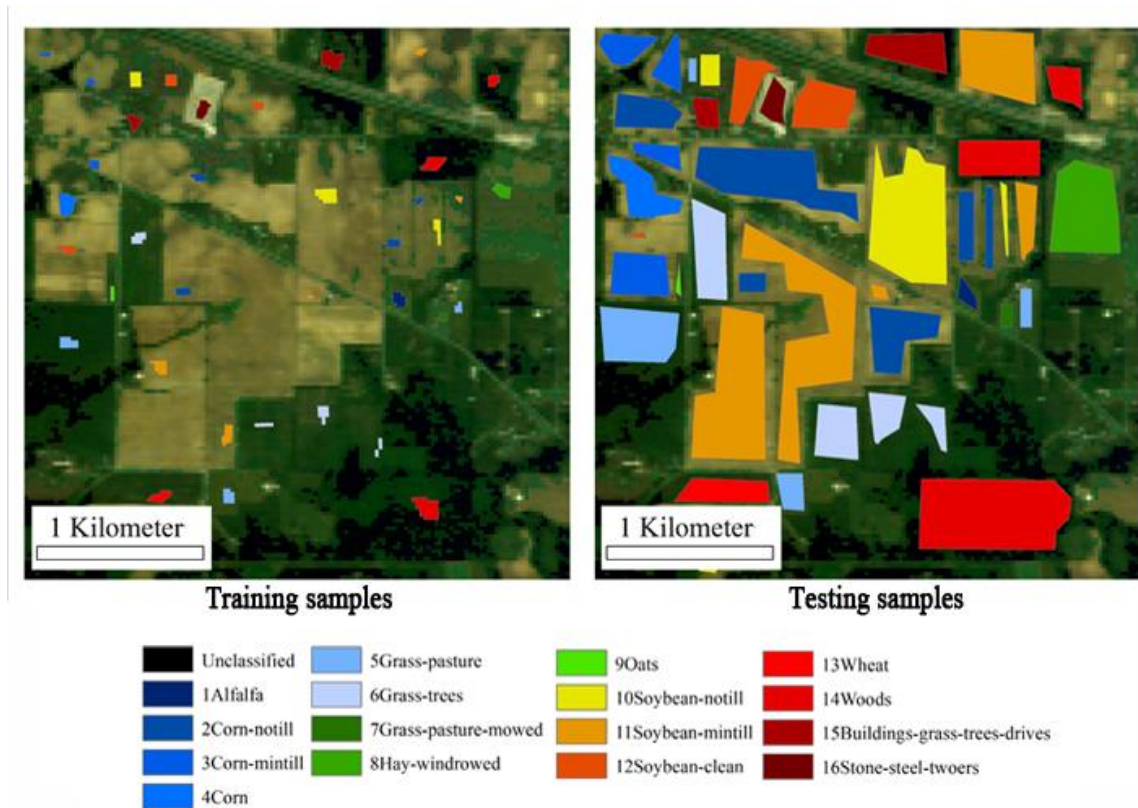


**Figure 4-14: Training and testing samples (AVIRIS image).**

Figure 4-15 shows the classification results of the original image and the best classification results from the 15 PCA classification maps for the Surrey, BC, CASI image. Figure 4-16 shows the classification results of the original image and the best classification results from the 15 PCA classification maps for the Indian Pines, AVIRIS image. The overall accuracies of the original images are 79.1% for the Surrey, CASI image and 80.1% for the Indian Pines, AVIRIS image. The best overall accuracies for the PCA images are 79.0% for the Surrey, BC, CASI image and 83.4% for the Indian Pines, AVIRIS image. In the classification results of the Surrey, BC, CASI image, the most noticeable error happens in shadow areas in Circle A and B, where trees are misclassified as dark shingle. The other noticeable error is that the painted asphalt areas are misclassified as concrete, with case in Circle C. In the classification results of the Indian Pines, AVIRIS image, both the original- and PCA-derived classification maps have small misclassified areas within most croplands. Yet, the PCA works better in several crop lands in Circle A, B, and C.
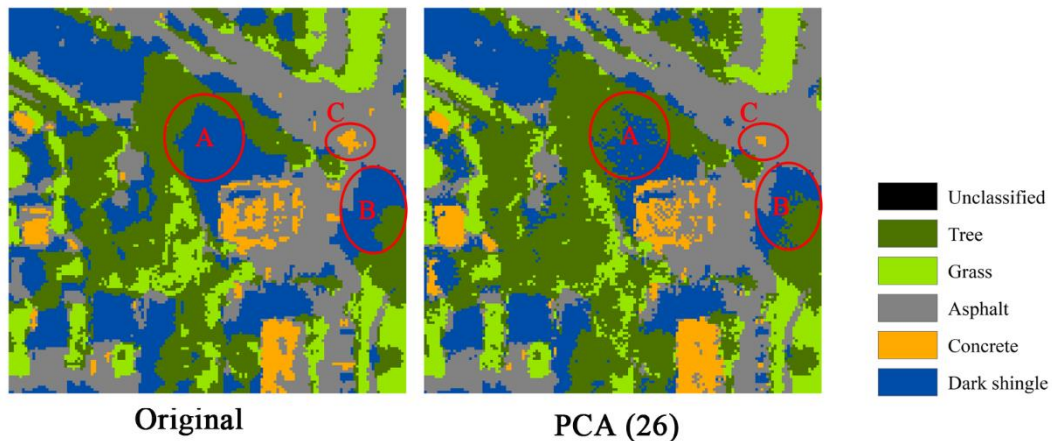


**Figure 4-15: Original image and PCA (preserved dimensions) image classification (CASI image).**
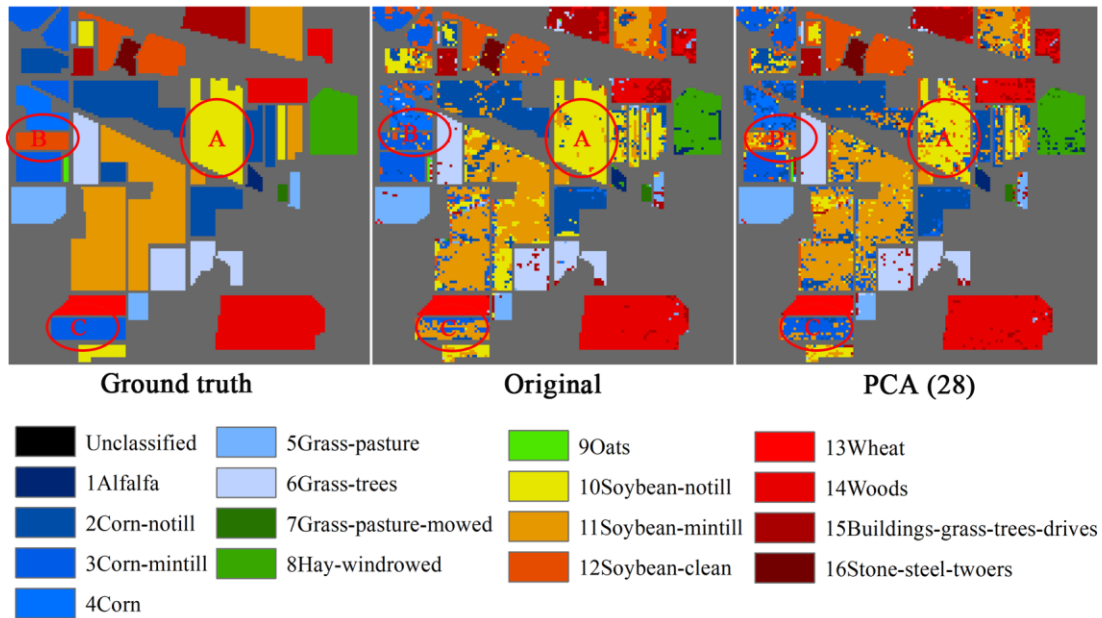
**Figure 4-16: Ground truth, original image classification, and PCA (preserved dimensions) image classification (AVIRIS image).**

## 4.4.3.1    Patch-based dimension reduction methods

For the 38 dimension-reduced images derived for each studied image using the patch-based methods, each will be classified 15 times by the SVM classifier using different amounts of preserved bands: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, and 30. In total, we have classification results for $38 \times 15$ situations for each of the studied images. The overall accuracy is the most common standard for comparing classification performances. It is calculated by dividing the number of pixels that are correctly classified by the total number of pixels. Regarding the eight different LPP- and NPE-based DR methods, the highest overall accuracies among different window sizes are shown in Table 4-1. It can be observed that the LPP and NPE versions of DR methods, separately targeting at preserving distance between pixels and preserving reconstructing weights between pixels and their neighbors, have similar overall accuracies. For both of the studied hyperspectral images, the LPP and NPE methods are able to pack the useful information within fewer bands and provide a higher overall accuracy than the traditional PCA method. With proper window size, the tested three patch-based methods (IPD, RCM,

and WRCM) boost the overall accuracy from the traditional LPP and NPE methods. The IPD-based LPP and NPE methods tend to achieve better classification accuracy with smaller window sizes, and contrarily, the RCM- and WRCM-based LPP and NPE methods tend to achieve better classification accuracy with larger window sizes. In most cases, the WRCM-based LPP and NPE methods outperform the other patch-based methods.

**Table 4-1: Highest overall accuracy among patch-based dimension reduction methods.**

| Surrey, BC, CASI Method | OA | Indian Pines, AVIRIS Method | OA |
|---|---|---|---|
| LPP | 79.56% | LPP | 79.7% |
| LPP+IPD (3) | 80.15% | LPP+IPD (3) | 77.8% |
| LPP+RCM (7) | 87.27% | LPP+RCM (13) | 85.6% |
| LPP+WRCM (7) | **87.32%** | LPP+WRCM (13) | **85.7%** |
| NPE | 86.32% | NPE | 73.6% |
| NPE+IPD (9) | 85.34% | NPE+IPD (3) | 80.0% |
| NPE+RCM (11) | 86.84% | NPE+RCM (11) | **86.7%** |
| NPE+WRCM (13) | **87.11%** | NPE+WRCM (11) | 86.6% |

*OA: overall accuracy.
*Number in bold font indicates the highest accuracies separately among LPP-, NPE-, TLPP-, and TNPE-based methods.
*The number in each parenthesis after the method is the corresponding window size for the highest overall accuracy.

The best classification results from each LPP- and NPE-based method are shown in Figure 4-17 for Surrey, BC, CASI image and in Figure 4-18 for Indian Pines, AVIRIS image. Compared to the classification results from the original and PCA-derived images in Figure 4-15, the tree area in shadow (circle A) in Figure 4-15 is better correctly classified and the painted asphalt areas are less misclassified as concrete (circle C). Yet, the tree areas in shadow (circle B) still suffer from misclassification. In the Surrey, BC, CASI image classification maps (Figure 4-17), the LPP-based methods appear to be more accurate in determining concrete roof (circle D). Contrarily, the NPE-based methods appear to be more accurate in determining dark shingle roof (circle E). Further, LPP-based methods tend to misclassify tree areas as grass areas (circle F). Considering the three different patch-based adjacency graph/weight matrix calculation methods, the IPD-based method provides the worst results in both LPP- and NPE-based DR. In the Indian

Pines, AVIRIS image classification maps (Figure 4-18), the small portion of misclassified areas were eliminated, compared to the classification maps in Figure 4-16.
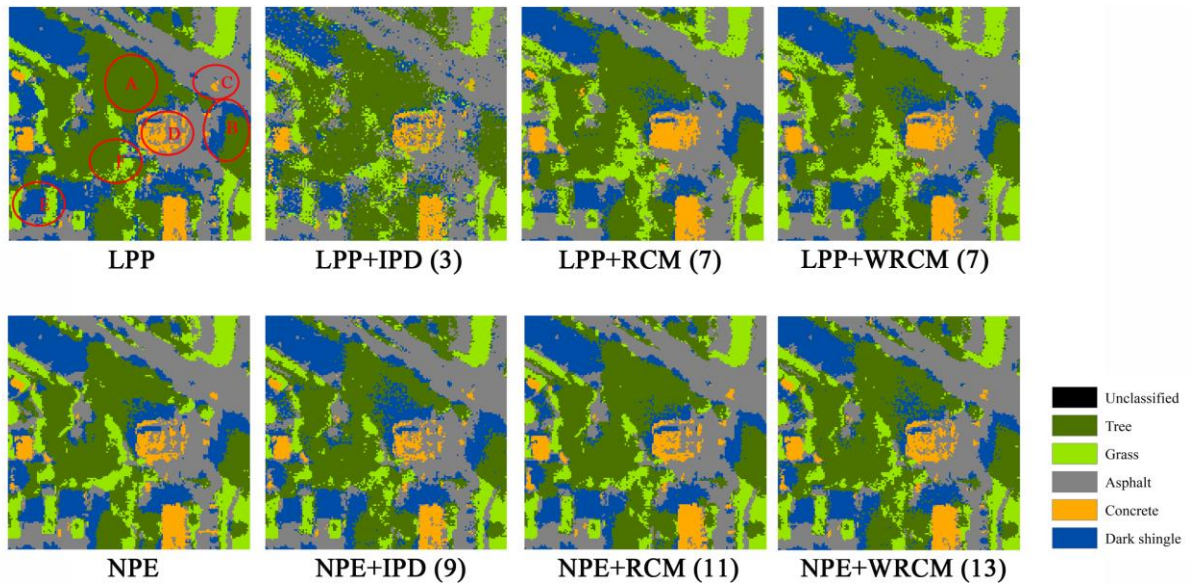


**Figure 4-17: Land cover classification maps derived from patch-based dimension-reduced images (CASI image).**
**(The number in each parenthesis after the method is the corresponding window size for the highest overall accuracy.)**
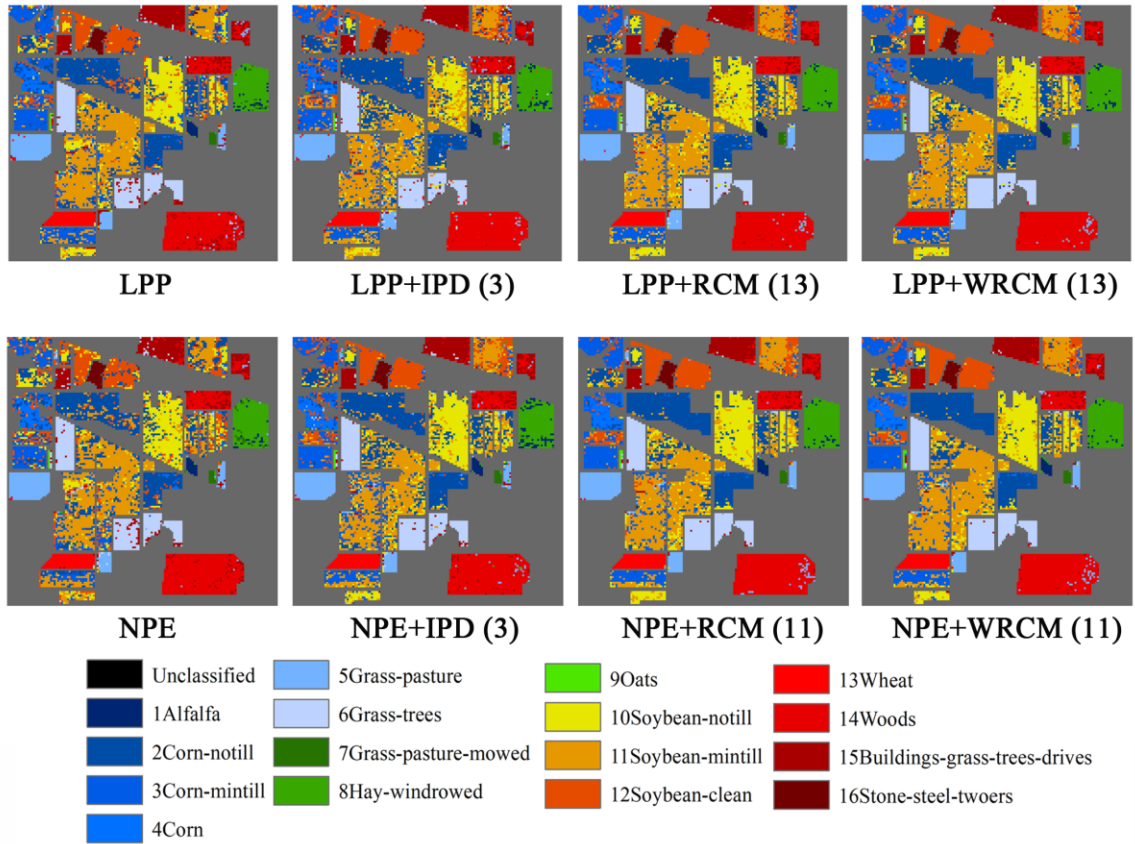
**Figure 4-18: Classification maps for patch-based DR results (AVIRIS image) (The number in each parenthesis after the method is the corresponding window size for the highest overall accuracy.)**

## 4.4.3.2    Tensor-patch-based dimension reduction methods

For the 48 dimension-reduced images derived for each studied image using the tensor-patch-based methods, each will be classified 15 times by the SVM classifier using different amounts of preserved bands: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, and 30. In total, we have classification results for $48 \times 15$ situations for each of the studied images. Regarding the eight different TLPP- and TNPE-based DR methods, the highest overall accuracies among different window sizes are shown in Table 4-2. It can be observed that the TLPP and TNPE versions of DR methods, separately targeted at preserving distance between pixels and preserving reconstructing weights between pixels and their neighbors, have similar overall accuracies. For both of the studied hyperspectral images, the TLPP and TNPE methods are able to achieve higher overall accuracies than the above LPP and NPE methods.  With proper window sizes, the tested three patch-

based methods (IPD, RCM, and WRCM) boost the overall accuracy from the traditional TLPP and TNPE methods. The IPD-based TLPP methods tend to achieve better classification accuracy with larger window sizes, and contrarily, IPD-based TNPE methods tend to achieve better classification accuracy with smaller window size. The RCM- and WRCM-based TLPP and TNPE methods tend to achieve great classification accuracy between window sizes of $7 \times 7$ to $11 \times 11$. In most cases, the WRCM-based TLPP and TNPE methods outperform the other patch-based methods.

**Table 4-2 Highest overall accuracy among tensor-patch-based DR methods.**

| Surrey, BC，CASI Method | OA | Indian Pines, AVIRIS Method | OA |
|---|---|---|---|
| TLPP (5) | 86.80% | TLPP (5) | 93.3% |
| TLPP+IPD(3) | 89.78% | TLPP+IPD(13) | 94.7% |
| TLPP+RCM(12) | 91.06% | TLPP+RCM(7) | 98.4% |
| TLPP+WRCM(12) | **91.27%** | TLPP+WRCM(9) | **99.5%** |
| TNPE (3) | 84.00% | TNPE (5) | 95.1% |
| TNPE+IPD(5) | 89.23% | TNPE+IPD(3) | 95.0% |
| TNPE+RCM (7) | **90.67%** | TNPE+RCM (11) | 98.3% |
| TNPE+WRCM(9) | 90.58% | TNPE+WRCM(11) | **98.4%** |

*OA: overall accuracy.
*Number in bold font indicates the highest accuracies separately among LPP-, NPE-, TLPP-, and TNPE-based methods.
*The number in each parenthesis after the method is the corresponding window size for the highest overall accuracy.

The best classification results from each method are shown in Figure 4-19 for Surrey, BC, CASI image and in Figure 4-20 for Indian Pines, AVIRIS image. Compared to the above classification results from patch-based methods, the classification results from tensor-patch-based methods suffer much less from small misclassified areas for both of the studied images. In Figure 4-19, the problem of misclassified tree under shadow that appears at different levels in the results of PCA, LPP, and NPE methods is greatly alleviated (circle A and B). The misclassification of painted asphalt is alleviated in TLPP, TNPE, and their IPD versions, and is completely avoided in RCM- and WRCM-based TLPP and TNPE methods (circle C). The misclassifications of concrete and dark shingle roofs (circle D and E) and grass (circle F) are also eliminated in the tensor-patch-based

methods, compared to patch-based methods. In Figure 4-20, the small misclassified areas within most of the croplands are avoided. In circle A and D, TNPE-based results show better accuracy than the TLPP-based method.
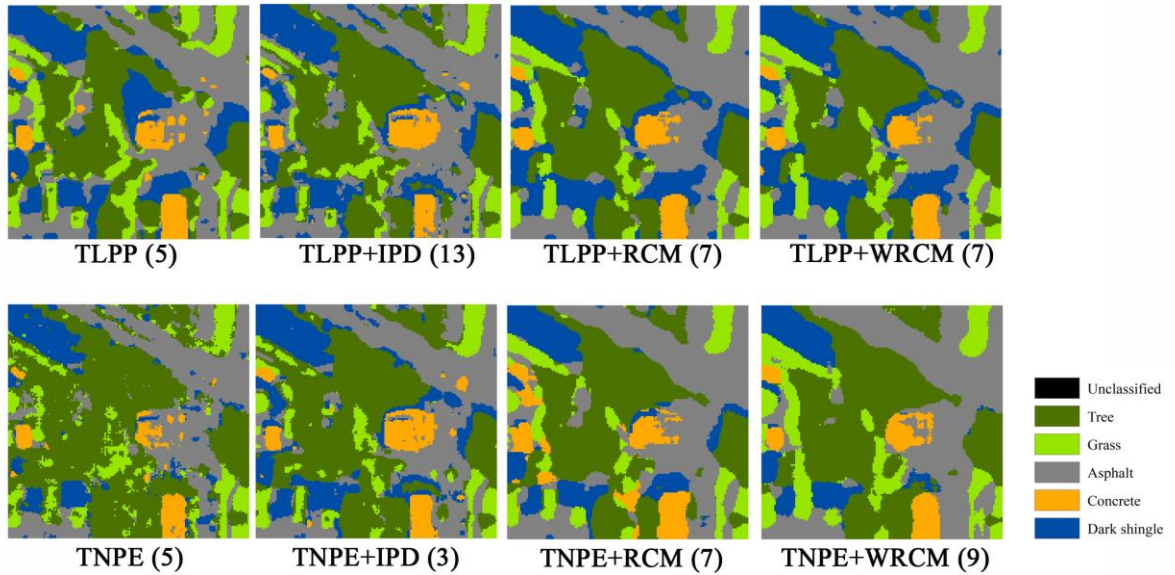


**Figure 4-19: Classification maps for tensor-patch-based DR results (CASI image).**
**(The number in each parenthesis after the method is the corresponding window size**
**for the highest overall accuracy.)**
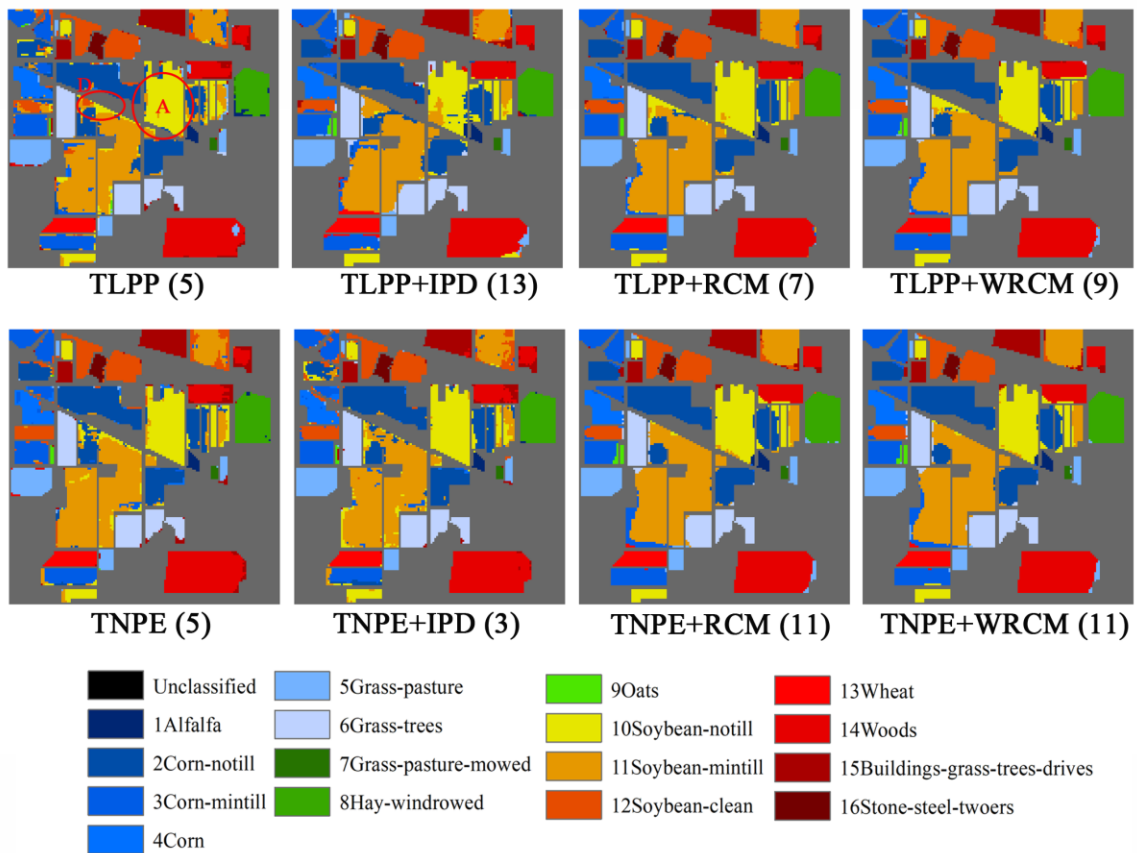
**Figure 4-20: Classification maps for tensor-patch-based DR (AVIRIS image).**
**(The number in each parenthesis after the method is the corresponding window size**
**for the highest overall accuracy.)**

## 4.5 Discussion

In this section, we will separately discuss the three contributions of this research listed in 4.1.

### 4.5.1 Comparison between patch-based and tensor-patch based dimension reduction methods

From the classification results, we can deduce that both patch-based and tensor-patch-based DR methods increase the overall accuracy from the traditional methods by including spatial information in process. In the case of the Surrey, BC, CASI image, the tensor-patch-based methods achieve ~5% higher overall accuracy than the patch-based

methods. In the case of the Indian Pines, AVIRIS image, the tensor-patch-based methods achieve ~10% higher overall accuracy than the patch-based methods. This is because the tensor-patch-based methods incorporate more spatial information in DR than the patch-based method. The Surrey, BC, CASI image of urban scene has more irregular shapes and more abrupt changes in land cover types than the Indian Pines, AVIRIS image of the agricultural scene. In other words, the Indian Pines, AVIRIS image has more different local spatial patterns between different land cover types than the Surrey, BC, CASI image. Thus, the Indian Pines, AVIRIS image benefited more from the preserved spatial information than the Surrey, BC, CASI image. Furthermore, the Surrey, BC, CASI image has a much higher spatial resolution of 1 m than the 20 m spatial resolution of the Indian Pines, AVIRIS image. The coarse spatial resolution of the Indian Pines, AVIRIS image may also contribute to the better performance of the tensor-patch-based methods. The coarse spatial resolution leads to smoother land cover change, which decreases the local spatial distinction between the same land cover and increase the local spatial distinction between different land cover. Another finding relates to the different objectives of the LPP/TLPP and NPE/TNPE, which separately target at preserving distance between pixels and preserving reconstructing weights between pixels and their neighbors. The LPP/TLPP methods provides better results in the Surrey, BC, CASI image, while the NPE/TNPE methods provides better results in the Indian Pines, AVIRIS image. In addition, The IPD-derived adjacency graph/weight matrix works better with the tensor-patch-based methods than with the patch-based methods for both of the studied images.

## 4.5.2    WRCM

From the Table 4-2 and 4-3 above, it can be observed that WRCM-based patch/tensor-patch methods often achieve the best overall accuracy among the three different patch-based adjacency graph/weight matrix calculation methods. In addition, the individual producer's and user's accuracy of each class were calculated for the classification results of the highest overall accuracies within the LPP-, NPE-, TLPP-, and TNPE-group of methods. The producer's accuracy reflects the omission error when a class *A* pixel fails to be classified as class *A*, and the user's accuracy reflects the commission error when a class *B* pixel is accidentally classified as class *A*. Table 4-3, 4-4, 4-5, and 4-6 separately

show the producer's and user's accuracies for the Surrey, BC, CASI image. Table 4-7, 4-8, 4-9, and 4-10 separately show the producer's and user's accuracies for the Indian Pines, AVIRIS image. In the case of the Surrey, BC, CASI image (Table 4-3, 4-4, 4-5, and 4-6), the WRCM-based method does not have obvious advantages over the IPD- and RCM-based methods. Some common errors for most of the methods include: (1) both omission and commission errors in the class of grass; (2) omission errors in the class of concrete; and (3) commission error in the class of dark shingle. On the other hand, in the case of the Indian Pines, AVIRIS image (Table 4-7, 4-8, 4-9, and 4-10), the WRCM-based method shows overwhelming superiority over the IPD- and RCM-based methods. This may be due to the different spatial pattern found in the urban and agricultural environments. Also, agreeing with the overall accuracy, the NPE/TNPE methods have a higher producer's and user's accuracy in the Indian Pines, AVIRIS image.

**Table 4-3: Producer's accuracies for patch-based DR methods (CASI image).**

| Producer's accuracy (%) | PCA | LPP | LPP IPD | LPP RCM | LPP WRCM | NPE | NPE IPD | NPE RCM | NPE WRCM |
|---|---|---|---|---|---|---|---|---|---|
| Tree | 84.72 | 91.39 | 87.22 | 85.36 | 85.59 | 89.06 | 87.54 | 92.77 | **92.98** |
| Grass | 65.35 | 65.46 | 67.83 | 89.29 | 89.14 | 76.61 | 88.86 | 89.71 | **90.00** |
| Asphalt | 81.25 | 77.14 | 87.43 | 88.36 | 87.50 | **90.02** | 84.4 | 84.53 | 84.11 |
| Concrete | 65.03 | 66.80 | 62.43 | **90.95** | 90.23 | 66.89 | 68.36 | 73.96 | 77.11 |
| Dark shingle | 83.12 | 81.31 | 70.88 | 84.08 | 87.53 | 95.48 | **95.68** | 89.42 | 89.86 |

*Number in bold font indicates the highest accuracies separately among patch-based and tensor-patch-based methods.

**Table 4-4: User's accuracies for patch-based DR methods (CASI image).**

| User's accuracy (%) | PCA | LPP | LPP IPD | LPP RCM | LPP WRCM | NPE | NPE IPD | NPE RCM | NPE WRCM |
|---|---|---|---|---|---|---|---|---|---|
| Tree | 82.87 | 78.56 | 76.06 | 92.18 | 92.34 | 87.91 | **93.39** | 92.26 | 91.61 |
| Grass | 77.82 | 82.75 | 75.23 | 65.65 | 66.67 | 78.52 | **83.49** | 79.70 | 78.95 |
| Asphalt | 85.47 | 85.54 | 86.57 | 94.46 | **94.74** | 89.94 | 91.23 | 92.19 | 93.77 |
| Concrete | 88.88 | 91.20 | 90.88 | 88.34 | **91.63** | 92.40 | 87.5 | 84.24 | 85.65 |
| Dark shingle | 61.57 | 66.6 | 66.87 | 71.85 | 69.54 | **76.06** | 60.61 | 68.10 | 67.44 |

*Number in bold font indicates the highest accuracies separately among patch-based and tensor-patch-based methods.

**Table 4-5: Producer's accuracies for tensor-patch-based DR methods (CASI image).**

| Producer's accuracy (%) | TLPP | TLPP IPD | TLPP RCM | TLPP WRCM | TNPE | TNPE IPD | TNPE RCM | TNPE WRCM |
|---|---|---|---|---|---|---|---|---|
| Tree | 82.41 | 93.66 | 93.39 | 93.46 | 94.97 | **95.43** | 94.50 | 94.46 |
| Grass | 76.61 | 73.02 | **93.86** | 93.71 | 68.78 | 72.49 | 90.57 | 92.86 |
| Asphalt | 92.79 | **93.00** | 91.86 | 92.55 | 88.20 | 90.23 | 91.76 | 90.98 |
| Concrete | 74.76 | 89.18 | **95.52** | 94.81 | 61.86 | 91.65 | 86.88 | 92.47 |
| Dark shingle | **99.17** | 85.53 | 76.63 | 76.54 | 79.26 | 83.41 | 79.72 | 76.19 |

*Number in bold font indicates the highest accuracies separately among patch-based and tensor-patch-based methods.

**Table 4-6: User's accuracies for tensor-patch-based DR methods (CASI image).**

| User's accuracy (%) | TLPP | TLPP IPD | TLPP RCM | TLPP WRCM | TNPE | TNPE IPD | TNPE RCM | TNPE WRCM |
|---|---|---|---|---|---|---|---|---|
| Tree | 87.99 | 84.89 | **91.90** | 91.56 | 79.42 | 83.08 | 91.55 | 86.42 |
| Grass | 75.5 | 89.49 | 82.64 | 83.14 | 83.76 | 88.50 | **93.79** | 83.33 |
| Asphalt | 93.06 | 93.99 | 93.02 | 93.07 | 87.45 | **96.69** | 93.25 | 93.52 |
| Concrete | 95.98 | **96.21** | 93.90 | 94.91 | 97.90 | 91.48 | 89.89 | 95.68 |
| Dark shingle | 72.75 | 84.13 | 84.70 | 86.03 | 77.97 | 82.72 | 77.73 | **93.00** |

*Number in bold font indicates the highest accuracies separately among patch-based and tensor-patch-based methods.

**Table 4-7: Producer's accuracies for patch-based DR methods (AVIRIS image).**

| Producer's accuracy (%) | PCA | LPP | LPP IPD | LPP RCM | LPP WRCM | NPE | NPE IPD | NPE RCM | NPE WRCM |
|---|---|---|---|---|---|---|---|---|---|
| Alfalfa | 98.22 | 98.56 | 98.60 | 93.33 | 99.52 | 98.36 | 98.95 | 99.14 | **99.85** |
| Corn-notill | 80.91 | 75.74 | 80.41 | 77.97 | 85.99 | 66.9 | 82.34 | 85.48 | **86.29** |
| Corn-mintill | 76.37 | 75.9 | 66.2 | 69.64 | 79.34 | 52.74 | 69.33 | **83.57** | 81.06 |
| Corn | 66.05 | 64.2 | 57.41 | 62.35 | 70.37 | 65.43 | 70.37 | 78.4 | **82.72** |
| Grass-pasture | 92.89 | 91.05 | 90.53 | 94.47 | 95 | 92.63 | 93.95 | 97.11 | **96.58** |
| Grass-trees | 97.05 | 95.09 | 94.89 | 98.04 | 97.64 | 93.91 | 97.84 | 99.41 | **99.80** |
| Grass-pasture | 97.23 | 98.52 | 98.60 | 94.44 | 98.95 | 97.85 | 98.96 | 98.56 | **99.28** |
| Hay-windrowe | 90.13 | 89.62 | 90.63 | **97.72** | 94.94 | 89.11 | 85.32 | 95.70 | 93.42 |
| Oats | 97.19 | 98.52 | 98.66 | 88.89 | **99.62** | 98.56 | 98.62 | 98.93 | 99.18 |
| Soybean-notill | 77.35 | 74.4 | 55.76 | 67.29 | 78.82 | 68.23 | 77.08 | 84.05 | **85.66** |
| Soybean-mintill | 72.33 | 67.99 | 65.32 | 66.27 | 71.97 | 58.14 | 61.58 | 73.34 | **73.52** |
| Soybean-clean | 99.31 | 94.48 | 97.19 | 97.59 | 99.39 | 83.45 | 99.66 | 99.16 | **99.82** |
| Wheat | 98.16 | 98.22 | 98.49 | 99.33 | **99.80** | 98.65 | 98.45 | 98.39 | 99.68 |
| Woods | 90.51 | 81.92 | 86.87 | 89.39 | 90.71 | 87.07 | 87.37 | **91.21** | 89.9 |
| Buildings-g | 98.04 | 99.22 | 99.60 | 94.51 | 99.61 | 90.98 | 98.04 | 98.43 | **99.89** |
| Stone-steel | 98.56 | 98.93 | 99.08 | 98.18 | 98.59 | 98.44 | 98.93 | 99.10 | **99.89** |

*Number in bold font indicates the highest accuracies separately among patch-based and tensor-patch-based methods.

**Table 4-8: User's accuracies for patch-based DR methods (AVIRIS image).**

| User's accuracy (%) | PCA | LPP | LPP IPD | LPP RCM | LPP WRCM | NPE | NPE IPD | NPE RCM | NPE WRCM |
|---|---|---|---|---|---|---|---|---|---|
| Alfalfa | 53.57 | 58.82 | 49.18 | **77.78** | 60.00 | 61.22 | 35.71 | 63.83 | 54.55 |
| Corn-notill | 75.05 | 69.85 | 71.29 | 78.69 | 83.2 | 58.22 | 73.46 | 83.04 | **84.83** |
| Corn-mintill | 77.96 | 73.04 | 68.01 | 65.35 | 74.67 | 61.38 | 67.63 | **79.94** | 77.2 |
| Corn | **89.92** | 87.39 | 77.5 | 72.14 | 83.21 | 78.52 | 84.44 | 89.44 | 83.23 |
| Grass-pasture | **96.45** | 91.29 | 88.89 | 93.25 | 93.77 | 95.91 | 90.84 | 91.34 | 90.17 |
| Grass-trees | 96.86 | 98.57 | 93.97 | 97.27 | 98.61 | 94.09 | 98.03 | **98.83** | 98.64 |
| Grass-pasture | 58.06 | 48.65 | 85.71 | 94.44 | **99.86** | 42.86 | 81.82 | **99.25** | 94.74 |
| Hay-windrowe | 99.11 | 98.88 | 98.62 | 99.23 | 99.47 | 98.78 | 97.40 | 98.95 | **99.28** |
| Oats | 98.62 | 98.56 | 98.71 | 66.67 | **99.62** | 98.63 | 98.69 | 99.50 | 99.37 |
| Soybean-notill | 76.42 | 71.25 | 69.1 | 65.36 | 75.19 | 65.85 | 75.16 | 78.18 | **79.18** |
| Soybean-mintill | 85.00 | 81.61 | 72.18 | 76.02 | 83.13 | 73.5 | 79.71 | 85.53 | **86.76** |
| Soybean-clean | 58.06 | 59.44 | 59.55 | 61.52 | 67.44 | 42.01 | 54.22 | 68.08 | **67.92** |
| Wheat | 98.21 | 98.60 | 92.16 | 97.24 | 99.30 | 98.60 | 98.60 | 98.34 | **99.63** |
| Woods | 99.33 | 99.14 | 99.19 | 98.99 | 99.78 | 98.63 | 99.54 | 99.56 | **99.92** |
| Buildings-g | 68.49 | 56.47 | 66.06 | 69.25 | 73.84 | 59.49 | 68.31 | **78.19** | 77.04 |
| Stone-steel | 90.16 | 91.67 | 90.16 | 98.18 | **99.65** | 96.49 | 98.56 | 99.16 | 96.49 |

*Number in bold font indicates the highest accuracies separately among patch-based and tensor-patch-based methods.

**Table 4-9: Producer's accuracies for tensor-patch-based DR methods (AVIRIS image).**

| Producer's accuracy (%) | TLPP | TLPP IPD | TLPP RCM | TLPP WRCM | TNPE | TNPE IPD | TNPE RCM | TNPE WRCM |
|---|---|---|---|---|---|---|---|---|
| Alfalfa | 98.56 | 98.79 | 99.23 | 99.68 | 98.20 | 98.72 | 99.31 | **99.84** |
| Corn-notill | 91.47 | 94.92 | 99.70 | 98.27 | 97.16 | 93.60 | 98.17 | **99.95** |
| Corn-mintill | 89.83 | 94.21 | 99.53 | 99.06 | 94.68 | 88.42 | 99.06 | **99.84** |
| Corn | 99.38 | 99.29 | 99.35 | **99.86** | 91.36 | 97.53 | 98.41 | 99.66 |
| Grass-pasture | 97.37 | 98.79 | 98.82 | **99.83** | 98.95 | 98.95 | 98.79 | 99.76 |
| Grass-trees | 98.56 | 99.61 | 98.56 | 99.46 | 98.34 | 98.73 | 99.18 | **99.95** |
| Grass-pasture | 98.27 | 98.83 | 99.10 | 99.26 | 98.16 | 98.63 | 99.11 | **99.46** |
| Hay-windrowe | 95.19 | 98.48 | 99.24 | 99.75 | 99.24 | 98.69 | 99.24 | **99.78** |
| Oats | 98.52 | 98.86 | 99.14 | **99.92** | 98.36 | 98.49 | 99.21 | 99.69 |
| Soybean-notill | 93.16 | 86.6 | 98.26 | 99.06 | 94.37 | 98.26 | 98.79 | **99.65** |
| Soybean-mintill | 87.23 | 89.96 | 95.55 | 95.72 | 88.84 | 90.08 | 95.31 | **99.82** |
| Soybean-clean | 98.54 | 98.93 | 98.96 | 99.43 | 98.54 | 98.72 | 99.10 | **99.96** |
| Wheat | 98.05 | 98.41 | 98.72 | 99.38 | 98.17 | 98.50 | 99.24 | **99.66** |
| Woods | 95.25 | 97.58 | 97.78 | 98.59 | 95.56 | 96.06 | 98.79 | **99.78** |
| Buildings-g | 99.26 | 99.42 | 99.68 | 99.80 | 99.08 | 99.61 | 99.71 | **99.86** |
| Stone-steel | 98.63 | 98.46 | 99.70 | 99.54 | 98.52 | 98.97 | 99.73 | **99.93** |

*Number in bold font indicates the highest accuracies separately among patch-based and tensor-patch-based methods.

**Table 4-10: User's accuracies for tensor-patch-based DR methods (AVIRIS image).**

| User's accuracy (%) | TLPP | TLPP IPD | TLPP RCM | TLPP WRCM | TNPE | TNPE IPD | TNPE RCM | TNPE WRCM |
|---|---|---|---|---|---|---|---|---|
| Alfalfa | 61.22 | 83.33 | 98.38 | 99.49 | 90.91 | 98.64 | 98.73 | **99.64** |
| Corn-notill | 89.12 | 95.8 | 98.79 | 99.69 | 93.73 | 93.51 | 99.38 | **99.89** |
| Corn-mintill | 89.55 | 89.45 | 95.78 | 91.74 | 94.09 | 95.76 | 92.01 | **99.76** |
| Corn | 98.17 | 99.39 | 98.78 | **99.86** | 97.37 | 91.33 | 99.26 | 99.48 |
| Grass-pasture | 92.96 | 94.06 | 99.48 | 97.69 | 99.47 | 99.73 | 97.19 | **99.88** |
| Grass-trees | 99.22 | 98.64 | 99.29 | 99.80 | 98.81 | 99.07 | 99.80 | **99.96** |
| Grass-pasture | 98.55 | 98.39 | 98.07 | 99.46 | 98.36 | 99.13 | 98.76 | **99.78** |
| Hay-windrowe | 98.43 | 98.95 | 99.28 | 99.08 | 98.97 | 98.41 | 99.45 | **99.76** |
| Oats | 98.59 | 64.29 | 99.03 | 69.23 | 98.33 | 99.19 | 69.23 | **99.61** |
| Soybean-notill | 90.49 | 92.95 | 96.45 | 98.14 | 93.74 | 89.94 | 98.14 | **99.84** |
| Soybean-mintill | 93.51 | 95.10 | 99.26 | 99.44 | 96.27 | 97.49 | 99.44 | **99.96** |
| Soybean-clean | 82.86 | 86.05 | 96.03 | 98.98 | 75.92 | 74.55 | 96.67 | **99.73** |
| Wheat | 98.60 | 83.93 | 98.27 | 98.60 | 98.04 | 98.06 | 98.60 | **99.54** |
| Woods | 99.79 | 99.90 | 98.07 | 99.58 | 99.19 | 99.06 | 98.74 | **99.92** |
| Buildings-g | 91.40 | 89.16 | 92.39 | 94.44 | 84.16 | 86.69 | 94.44 | **99.82** |
| Stone-steel | 98.64 | 98.43 | 99.06 | **99.94** | 98.36 | 98.53 | 99.78 | 99.46 |

*Number in bold font indicates the highest accuracies separately among patch-based and tensor-patch-based methods.

## 4.5.3 Computation complexity

In regards to the computation complexity of the applied DR methods, for a $k$-order tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_k}$ the patch-based method has a space complexity of $O(d^2)$ and time complexity of $O(d^3)$, where d is the total dimension of data: $d = \prod_{i=1}^{k} I_i$. The tensor-patch-based method has a space complexity of $O(nd)$ and time complexity of $O\left(k(n^2 d I_{max} + I_{max}^3)\right)$, where d is the total dimension of data, n is number of data points, and $I_{max} = \max(I_1, I_2, \dots, I_k)$. In the computer vision studies, the total dimensions of a dataset ($d$) are usually much larger than the number of data points ($n$), when the tensor-patch-based method requires less computation complexity than the patch-based method. However, in the case of a hyperspectral image, the number of data points ($n$) is usually much larger than the total dimensions of a dataset ($d$), which results in heavy computation complexity of the tensor-patch-based method. Table 4-11 shows the real processing time in this research. The processing times of the TLPP and TNPE methods is much longer than the LPP and NPE methods.

**Table 4-11: Time consumption.**

| Surrey, BC Method | Time (s) | Indian Pines Method | Time (s) |
|---|---|---|---|
| LPP | $2.4 \pm 0.3$ | LPP | $1.6 \pm 0.2$ |
| NPE | $0.2 \pm 0.1$ | NPE | $0.4 \pm 0.1$ |
| TLPP | $160.7 \pm 144.6$ | TLPP | $144.4 \pm 100.3$ |
| TNPE | $170.1 \pm 156.3$ | TNPE | $142.5 \pm 101.3$ |

### 4.5.4    Other observations

Two other issues concerning the Surrey, BC, CASI image are listed here. Firstly, one source of error is the shadows caused by tall objects, especially tall buildings. The two causes are as follows: (1) areas in shadow normally have low reflection and the limited information may lead to errors in classification in the hyperspectral imagery; (2) areas in shadow in the orthophoto (ground truth) may cause wrong judgement in the accuracy report. It is found that the tensor-patch-based methods best handled the shadow areas. Another unavoidable complication is tree cover over impervious surfaces, which often occurs along roads or beside residential buildings. As the Surrey airborne data only captures imageries from top-down, the impervious surface area that is covered by tree canopy is difficult to correctly classify.

## 4.6  Conclusion

This research performed two genres of spectral-spatial DR methods: patch-based and tensor-patch-based methods for two hyperspectral images with different spatial resolutions separately in urban and agricultural scenes. The patch-based methods employ the spatial information in the process of generating adjacency graph/weight matrix. The tensor-patch-based methods employ the spatial information in the processes of generating adjacency graph/weight matrix and solving the target eigenproblem. The three research questions in 4.2 were correspondingly addressed below.

(1) Both the patch-based and tensor-patch-based DR methods are able to pack the useful information within fewer bands and provide a higher overall accuracy than the traditional PCA method.

(2) The tensor-patch-based DR methods by preserving more spatial information outperform the patch-based DR methods in the resulting land cover classification. This advantage of the tensor-patch-based method is more noticeable for images with distinct local spatial patterns, meaning low local spatial differences between same land cover and high local spatial differences between different land cover. Images with block land cover patterns and coarse spatial resolutions are more likely to have distinct local spatial patterns. However this advantage of tensor-patch-based methods comes with intense computation complexity.

(3) The appropriate window sizes for the patch-based and tensor-patch-based methods vary regarding the specific method (LPP, NPE, TLPP, or TNPE) and the target image. We found that in general the IPD-based methods usually achieve better overall accuracy with smaller window sizes, and the RCM- and WRCM-based methods usually require larger window sizes to achieve better overall accuracy. The appropriate window sizes for the patch-based methods are usually larger than the appropriate window sizes for the tensor-patch-based methods.

(4) The proposed WRCM adjacency graph/weight matrix calculation method slightly outperforms the IPD and RCM methods, as it comprehensively depicts the local spatial distribution.

(5) The land cover classification results derived from this study are as good as or better than the results from up-to-date land cover mapping literatures (Deng et al. 2018, Ren et al. 2017, Pelletier et al. 2016). The tensor-based DR methods have been proved to be promising and valuable.

## References

Belkin, M. & P. Niyogi (2003) Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation,* 15**,** 1373-1396.

Chen, G. & S.-E. Qian (2007) Dimensionality reduction of hyperspectral imagery. *Journal of Applied Remote Sensing,* 1**,** 013509.

Cheriyadat, A. & L. M. Bruce. 2003. Why principal component analysis is not an appropriate feature extraction method for hyperspectral data. In *Geoscience and Remote Sensing Symposium, 2003. IGARSS'03. Proceedings. 2003 IEEE International*, 3420-3422. IEEE.

Cui, M., S. Prasad, W. Li & L. M. Bruce (2013) Locality preserving genetic algorithms for spatial-spectral hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 6**,** 1688-1697.

Dai, G. & D.-Y. Yeung. 2006. Tensor embedding methods. In *AAAI*, 330-335.

Daniel, S. F., A. Connolly, J. Schneider, J. VanderPlas & L. Xiong (2011) Classification of stellar spectra with local linear embedding. *The Astronomical Journal,* 142**,** 203.

Deng, Y.-J., H.-C. Li, L. Pan, L.-Y. Shao, Q. Du & W. J. Emery (2018) Modified Tensor Locality Preserving Projection for Dimensionality Reduction of Hyperspectral Images. *IEEE Geoscience and Remote Sensing Letters*.

Du, B., M. Zhang, L. Zhang, R. Hu & D. Tao (2017) PLTD: Patch-based low-rank tensor decomposition for hyperspectral images. *IEEE Transactions on Multimedia,* 19**,** 67-79.

Gao, Y., X. Wang, Y. Cheng & Z. J. Wang (2015) Dimensionality reduction for hyperspectral data based on class-aware tensor neighborhood graph and patch alignment. *IEEE transactions on neural networks and learning systems,* 26**,** 1582-1593.

Han, T. & D. G. Goodenough. 2005. Nonlinear feature extraction of hyperspectral data based on locally linear embedding (LLE). In *Geoscience and Remote Sensing Symposium, 2005. IGARSS'05. Proceedings. 2005 IEEE International*, 1237-1240. IEEE.

He, X., D. Cai & P. Niyogi. 2006. Tensor subspace analysis. In *Advances in neural information processing systems*, 499-506.

He, X., D. Cai, S. Yan & H.-J. Zhang. 2005. Neighborhood preserving embedding. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, 1208-1213. IEEE.

He, X. & P. Niyogi. 2004. Locality preserving projections. In *Advances in neural information processing systems*, 153-160.

Hotelling, H. (1933) Analysis of a complex of statistical variables into principal components. *Journal of educational psychology,* 24**,** 417.

Hu, D., G. Feng & Z. Zhou (2007) Two-dimensional locality preserving projections (2DLPP) with its application to palmprint recognition. *Pattern recognition,* 40**,** 339-342.

Karami, A., M. Yazdi & G. Mercier (2012) Compression of hyperspectral images using discrete wavelet transform and tucker decomposition. *IEEE journal of selected topics in applied earth observations and remote sensing,* 5**,** 444-450.

Khodr, J. & R. Younes. 2011. Dimensionality reduction on hyperspectral images: a comparative review based on artificial datas. In *Image and Signal Processing (CISP), 2011 4th International Congress on*, 1875-1883. IEEE.

Lee, J. A. & M. Verleysen. 2007. *Nonlinear dimensionality reduction*. Springer Science & Business Media.

--- (2012) Graph-based dimensionality reduction.

Lewandowski, M., J. Martinez-del-Rincon, D. Makris & J.-C. Nebel. 2010. Temporal extension of laplacian eigenmaps for unsupervised dimensionality reduction of time series. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, 161-164. IEEE.

Li, C., Y. Ma, J. Huang, X. Mei & J. Ma (2015) Hyperspectral image denoising using the robust low-rank tensor recovery. *JOSA A,* 32**,** 1604-1612.

Li, W., S. Prasad, J. E. Fowler & L. M. Bruce (2011) Locality-preserving discriminant analysis in kernel-induced feature spaces for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters,* 8**,** 894-898.

--- (2012) Locality-preserving dimensionality reduction and classification for hyperspectral image analysis. *IEEE Transactions on Geoscience and Remote Sensing,* 50**,** 1185-1198.

Lu, H., K. N. Plataniotis & A. N. Venetsanopoulos (2008) MPCA: Multilinear principal component analysis of tensor objects. *IEEE transactions on Neural Networks,* 19**,** 18-39.

Luo, Y., D. Tao, K. Ramamohanarao, C. Xu & Y. Wen (2015) Tensor canonical correlation analysis for multi-view dimension reduction. *IEEE transactions on Knowledge and Data Engineering,* 27**,** 3111-3124.

Ma, L., M. M. Crawford & J. Tian (2010) Anomaly detection for hyperspectral images based on robust locally linear embedding. *Journal of Infrared, Millimeter, and Terahertz Waves,* 31**,** 753-762.

Mohan, A., G. Sapiro & E. Bosch (2007) Spatially coherent nonlinear dimensionality reduction and segmentation of hyperspectral images. *IEEE Geoscience and Remote Sensing Letters,* 4**,** 206-210.

Pang, Y., Y. Yuan & X. Li (2008) Gabor-based region covariance matrices for face recognition. *IEEE Transactions on Circuits and Systems for Video Technology,* 18**,** 989-993.

Pearson, K. (1901) LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science,* 2**,** 559-572.

Pelletier, C., S. Valero, J. Inglada, N. Champion & G. Dedieu (2016) Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sensing of Environment,* 187**,** 156-168.

Pu, H., Z. Chen, B. Wang & G.-M. Jiang (2014) A novel spatial–spectral similarity measure for dimensionality reduction and classification of hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing,* 52**,** 7008-7022.

Qin, H., L. Qin, L. Xue & Y. Li (2012) A kernel Gabor-based weighted region covariance matrix for face recognition. *Sensors,* 12**,** 7410-7422.

Ren, Y., L. Liao, S. J. Maybank, Y. Zhang & X. Liu (2017) Hyperspectral image spectral-spatial feature extraction via tensor principal component analysis. *IEEE Geoscience and Remote Sensing Letters,* 14**,** 1431-1435.

Roweis, S. T. & L. K. Saul (2000) Nonlinear dimensionality reduction by locally linear embedding. *science,* 290**,** 2323-2326.

Schölkopf, B., A. Smola & K.-R. Müller (1998) Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation,* 10**,** 1299-1319.

Shawe-Taylor, J. & N. Cristianini. 2004. *Kernel methods for pattern analysis*. Cambridge university press.

Tenenbaum, J. B., V. De Silva & J. C. Langford (2000) A global geometric framework for nonlinear dimensionality reduction. *science,* 290**,** 2319-2323.

Teng, X., B. Wu, W. Yu & C. Liu (2005) A hand gesture recognition system based on local linear embedding. *Journal of Visual Languages & Computing,* 16**,** 442-454.

Torgerson, W. S. (1952) Multidimensional scaling: I. Theory and method. *Psychometrika,* 17**,** 401-419.

Tuzel, O., F. Porikli & P. Meer. 2006. Region covariance: A fast descriptor for detection and classification. In *European conference on computer vision*, 589-600. Springer.

Van Der Maaten, L., E. Postma & J. Van den Herik (2009) Dimensionality reduction: a comparative. *J Mach Learn Res,* 10**,** 66-71.

Velasco-Forero, S. & J. Angulo (2013) Classification of hyperspectral images by tensor modeling and additive morphological decomposition. *Pattern Recognition,* 46**,** 566-577.

Wang, Y. & Y. Wu (2010) Complete neighborhood preserving embedding for face recognition. *Pattern Recognition,* 43**,** 1008-1015.

Wang, Z. & B. He. 2011. Locality perserving projections algorithm for hyperspectral image dimensionality reduction. In *Geoinformatics, 2011 19th International Conference on*, 1-4. IEEE.

Xu, D., S. Yan, L. Zhang, S. Lin, H.-J. Zhang & T. S. Huang (2008) Reconstruction and recognition of tensor-based objects with concurrent subspaces analysis. *IEEE Transactions on Circuits and Systems for Video Technology,* 18**,** 36-47.

Zhang, L., L. Zhang, D. Tao & X. Huang (2013) Tensor discriminative locality alignment for hyperspectral image spectral–spatial feature extraction. *IEEE Transactions on Geoscience and Remote Sensing,* 51**,** 242-256.

Zhang, T., D. Tao, X. Li & J. Yang (2009) Patch alignment for dimensionality reduction. *IEEE Transactions on Knowledge and Data Engineering,* 21**,** 1299-1313.

Zhao, W. & S. Du (2016) Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. *IEEE Transactions on Geoscience and Remote Sensing,* 54**,** 4544-4554.

Zhong, Z., B. Fan, J. Duan, L. Wang, K. Ding, S. Xiang & C. Pan (2015) Discriminant tensor spectral–spatial feature extraction for hyperspectral image classification. *IEEE Geoscience and Remote Sensing Letters,* 12**,** 1028-1032.

Zhou, Y., J. Peng & C. P. Chen (2015) Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing,* 53**,** 1082-1095.

# 5 General discussion and conclusions

## 5.1 Summary

The distribution of urban impervious surface area (ISA) and land cover are important inputs in a wide range of urban physical and social studies. Chapters 2-4 of this thesis investigated methods to improve current methods for urban ISA/land cover mapping.

Chapter 2 addresses the application of the future EnMAP image in ISA mapping. The future launch of the German hyperspectral satellite EnMAP (Environmental mapping and Analysis Program) in 2019 will provide new opportunities for timely and global ISA mapping. The previously proposed EnMAP applications heavily relied on existing reference endmembers, which may be impractical on a global scale. To overcome this defect, the use of the nonnegative matrix factorization (NMF) method was suggested to extract the endmember directly from the EnMAP imagery. The recent NMF-based hyperspectral spectral unmixing no longer depends on pure pixels in the original image, like the traditional methods. The NMF simultaneously calculates the endmember and abundance maps, by applying linear algebra to decompose the original hyperspectral image into an endmember matrix and an abundance matrix. However, despite the 10-year application of NMF in the hyperspectral spectral unmixing, researchers focused more on the extracted endmembers than the abundance maps. Very limited researches have been done to evaluate the abundance maps derived from the NMF methods, which was investigated in my thesis. Three traditional spectral unmixing methods (e.g. N-Findr, PPI, and ICA) and four NMF-based methods with three different constraints (e.g. sparseness, convex volume, and nonlinearity) were used to obtain the series of endmember sets, ISA abundance and classification maps. In the results, the traditional spectral unmixing methods encountered great difficulty when working with the EnMAP image, with correlation coefficients less than 0.2 in the linear regression models between predicted and reference ISA percentages, and less than 50% overall accuracy in ISA classification maps. The NMF-based methods outperformed the traditional spectral unmixing methods,

by achieving correlation coefficients ~0.7 in the linear regression models between predicted and reference ISA percentage, and over 85% overall accuracy in ISA classification maps. However, the NMF-derived abundance maps had great underestimations in the ISA. I found that it is necessary to apply linear transformations for the obtained ISA abundance maps. I also found that the constrained NMF provides no improvements in ISA mapping results compared to the traditional NMF in the case of the simulated EnMAP image. However, the constrained NMF greatly improved the ISA mapping results from the Hydice urban image of high spatial resolution.

Chapter 3 is a continuation of work from the Chapter 2. In Chapter 2, the newly available genre of hyperspectral spectral unmixing methods based on NMF was thoroughly studied. However, the NMF is based on linear algebra, which needs to first convert the hyperspectral data cube into a 2D matrix for further processing. Due to this conversion, the spatial information in the relative positions of the pixels is lost. With the wide interest in multilinear algebra, tensor-based hyperspectral image processing becomes possible. The tensor-based spectral unmixing was firstly realized using the matrix-vector nonnegative tensor factorization (MVNTF) in 2017 (Qian et al. 2017). However, in the 2017 paper, limited experiments were provided for the evaluation of the obtained abundance maps. By comparing the ISA mapping results from the NMF- and the MVNTF-based methods, this thesis found that the MVNTF has greater advantages when working with medium-spatial-resolution than with the high-spatial-resolution hyperspectral images. Similar to the NMF methods, the MVNTF-derived abundance maps had great underestimations in the ISA and proper linear transformations were performed. Under the construction of MVNTF spectral unmixing, this research proposed to integrate three additional constraints (sparseness, volume, and nonlinearity) to the MVNTF cost function. I found that the three constraints improved the ISA abundance results. The accuracies of the ISA classification map experienced an average 2% growth due to the added constraints. The constraints also shortened the processing time.

Chapter 4 deals with the problem known as the *curse of dimensionality*. The majority of current dimension reduction methods are restricted to the use of only the spectral

information, and the spatial information is not considered. In order to overcome this defect, two spectral-spatial representations: patch-based and tensor-patch-based, were studied in this paper. The two groups of methods at different processing steps incorporate both the spectral and spatial information. To date, the popularity of the two solutions is confined to computer vision studies and their applications in hyperspectral dimension reduction are limited. This thesis focused on both of the patch-based and tensor-patch-based variations of a group of dimension reduction methods called graph-based that learns the data structure from adjacency graphs/weight matrices. In total, the traditional PCA method, two traditional graph-based methods, six patch-based and eight tensor-patch-based methods were used to perform dimension reduction. The resulting dimension-reduced images were then classified. It was found that the patch-based and tensor-patch-based variations greatly boost the final classification results by 5%-15% from the traditional PCA method. The tensor-patch-based variations delivered better land cover classification results than the patch-based variations, but accompanied by much longer processing times. As graph-based methods heavily rely on the calculation of adjacency graphs/weight matrices, this paper proposed the use of a new method: weighted region covariance matrix, to produce the adjacency graphs/weight matrices. In the results, the newly proposed method can further improve the dimension reduction results in both the patch-based and tensor-patch-based methods.

## 5.2  Conclusions

This thesis presented methodologies, experiments, and discussions to answer the research questions raised in the Introduction to the thesis:

(1) As the EnMAP image has a medium spatial resolution of 30m, the pure pixels are very rare. Among the tested spectral unmixing matrix-based methods, the most suitable method is found to be a genre called nonnegative matrix factorization (NMF). This method searches for the endmembers outside the existed pixels in the image and simultaneously generates the abundance map. In addition, three variations of NMF were

153

tested and further improved the resulting ISA abundance and classification maps derived from the simulated EnMAP image.

(2) The recently proposed MVNTF was successful in implementing the use of the multilinear algebra in the NMF-based spectral unmixing method. Although the tensor-based spectral unmixing is able to consider the local spatial information from the image, it does not guarantee better ISA mapping results. It has been found that the tensor-based spectral unmixing methods may obtain more advantages when working with images of medium spatial resolutions.

(3) The three variations for the NMF methods were successfully adopted in the MVNTF methods with adjustments. Improvements in the resulting ISA abundance and classification were found.

(4) Through the patch-based strategy, the matrix-based dimension reduction method can learn local spatial information. Through the tensor-patch-based strategy, the matrix-based dimension reduction method can be converted to a tensor-based method. Both of the strategies were able to improve the resulting land cover classification.

(5) The above patch-based and tensor-patch-based dimension reduction methods both stem from a group of dimension reduction methods called the graph-based. As the graph-based rely on the intermediate results: adjacency graph and weight matrix. A new method called the weighted region covariance matrix (WRCM) was proposed to obtain the intermediate results to use with the patch-based and tensor-patch-based dimension reduction methods. The WRCM method improved the resulting land cover classification.

## 5.3 Contributions

ISA and land cover are fundamental data for a wide range of subjects in the physical and social sciences, as well as in municipalities for urban planning purposes (Weng and Quattrochi 2006). An up-to-date understanding of the urban ISA and land cover provides insight into concerns such as natural hazards, uncontrolled development, deteriorating environmental quality, loss of primary agricultural lands, destruction of important

wetlands, and loss of wildlife habitat (Anderson 1976). The growing needs for accurate ISA and land cover data from a wide range of fields on a global, national, and local level continues to this day. After the development of the camera technology, remote sensing has become a popular data source for ISA and land cover mapping (Avery, Berlin and Berlin 1977). This thesis provided research on both the applications of the new hyperspectral data and the new hyperspectral methods for improving the ISA and land cover mapping. The resulting ISA/land cover maps are as good as or better than the up-to-date studies. The improved accuracy will assist future urban studies. The growing academic literatures of ISA and land cover mapping should benefit from the methods and analyses presented.

## References

Anderson, J. R. 1976. *A land use and land cover classification system for use with remote sensor data*. US Government Printing Office.

Avery, T. E., G. L. Berlin & G. L. Berlin. 1977. *Interpretation of aerial photographs*. Burgess.

Qian, Y., F. Xiong, S. Zeng, J. Zhou & Y. Y. Tang (2017) Matrix-vector nonnegative tensor factorization for blind unmixing of hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing,* 55**,** 1776-1792.

Weng, Q. & D. A. Quattrochi. 2006. *Urban remote sensing*. CRC Press.

# Appendix A: Details of the CASI and simulated EnMAP spectral bands.

Table A-1: Details of the CASI spectral bands.

| Band number | Band center (nm) | Bandwidth (nm) | Band number | Band center (nm) | Bandwidth (nm) |
|---|---|---|---|---|---|
| 1 | 367.6 | 9.6 | 37 | 712.1 | 9.5 |
| 2 | 377.2 | 9.6 | 38 | 721.6 | 9.5 |
| 3 | 386.8 | 9.7 | 39 | 731.1 | 9.6 |
| 4 | 396.5 | 9.6 | 40 | 740.7 | 9.5 |
| 5 | 406.1 | 9.6 | 41 | 750.2 | 9.5 |
| 6 | 415.7 | 9.6 | 42 | 759.7 | 9.6 |
| 7 | 425.3 | 9.6 | 43 | 769.3 | 9.5 |
| 8 | 434.9 | 9.6 | 44 | 778.8 | 9.5 |
| 9 | 444.5 | 9.6 | 45 | 788.3 | 9.6 |
| 10 | 454.1 | 9.6 | 46 | 797.9 | 9.5 |
| 11 | 463.7 | 9.6 | 47 | 807.4 | 9.5 |
| 12 | 473.3 | 9.6 | 48 | 816.9 | 9.6 |
| 13 | 482.9 | 9.5 | 49 | 826.5 | 9.5 |
| 14 | 492.4 | 9.6 | 50 | 836 | 9.5 |
| 15 | 502 | 9.6 | 51 | 845.5 | 9.6 |
| 16 | 511.6 | 9.6 | 52 | 855.1 | 9.5 |
| 17 | 521.2 | 9.5 | 53 | 864.6 | 9.6 |
| 18 | 530.7 | 9.6 | 54 | 874.2 | 9.5 |
| 19 | 540.3 | 9.6 | 55 | 883.7 | 9.6 |
| 20 | 549.9 | 9.5 | 56 | 893.3 | 9.5 |
| 21 | 559.4 | 9.6 | 57 | 902.8 | 9.6 |
| 22 | 569 | 9.5 | 58 | 912.4 | 9.5 |
| 23 | 578.5 | 9.6 | 59 | 921.9 | 9.6 |
| 24 | 588.1 | 9.5 | 60 | 931.5 | 9.6 |
| 25 | 597.6 | 9.6 | 61 | 941.1 | 9.5 |
| 26 | 607.2 | 9.5 | 62 | 950.6 | 9.6 |
| 27 | 616.7 | 9.5 | 63 | 960.2 | 9.6 |
| 28 | 626.2 | 9.6 | 64 | 969.8 | 9.6 |
| 29 | 635.8 | 9.5 | 65 | 979.4 | 9.5 |
| 30 | 645.3 | 9.6 | 66 | 988.9 | 9.6 |
| 31 | 654.9 | 9.5 | 67 | 998.5 | 9.6 |
| 32 | 664.4 | 9.5 | 68 | 1008.1 | 9.6 |
| 33 | 673.9 | 9.6 | 69 | 1017.7 | 9.6 |
| 34 | 683.5 | 9.5 | 70 | 1027.3 | 9.6 |
| 35 | 693 | 9.5 | 71 | 1036.9 | 9.6 |
| 36 | 702.5 | 9.6 | 72 | 1046.5 | 9.6 |

Table A-2: Details of the simulated EnMAP spectral bands.

| Band number | Band center (nm) | Bandwidth (nm) | Band number | Band center (nm) | Bandwidth (nm) |
|---|---|---|---|---|---|
| 1 | 423 | 6.9 | 45 | 658 | 7.8 |
| 2 | 429 | 6.6 | 46 | 665 | 7.9 |
| 3 | 434 | 6.3 | 47 | 671 | 7.9 |
| 4 | 440 | 6.2 | 48 | 678 | 8 |
| 5 | 445 | 6 | 49 | 685 | 8.1 |
| 6 | 450 | 5.9 | 50 | 691 | 8.2 |
| 7 | 455 | 5.9 | 51 | 698 | 8.3 |
| 8 | 460 | 5.8 | 52 | 705 | 8.3 |
| 9 | 464 | 5.8 | 53 | 712 | 8.4 |
| 10 | 469 | 5.8 | 54 | 719 | 8.5 |
| 11 | 474 | 5.7 | 55 | 726 | 8.6 |
| 12 | 479 | 5.7 | 56 | 733 | 8.7 |
| 13 | 484 | 5.8 | 57 | 740 | 8.7 |
| 14 | 488 | 5.8 | 58 | 748 | 8.8 |
| 15 | 493 | 5.8 | 59 | 755 | 8.9 |
| 16 | 498 | 5.8 | 60 | 762 | 8.9 |
| 17 | 503 | 5.8 | 61 | 770 | 9 |
| 18 | 508 | 5.9 | 62 | 777 | 9.1 |
| 19 | 513 | 5.9 | 63 | 785 | 9.1 |
| 20 | 518 | 6 | 64 | 793 | 9.2 |
| 21 | 523 | 6 | 65 | 800 | 9.3 |
| 22 | 528 | 6.1 | 66 | 808 | 9.3 |
| 23 | 533 | 6.1 | 67 | 816 | 9.4 |
| 24 | 538 | 6.2 | 68 | 823 | 9.4 |
| 25 | 543 | 6.2 | 69 | 831 | 9.5 |
| 26 | 548 | 6.3 | 70 | 839 | 9.5 |
| 27 | 553 | 6.4 | 71 | 847 | 9.5 |
| 28 | 559 | 6.4 | 72 | 855 | 9.6 |
| 29 | 564 | 6.5 | 73 | 863 | 9.6 |
| 30 | 569 | 6.6 | 74 | 871 | 9.6 |
| 31 | 575 | 6.6 | 75 | 879 | 9.7 |
| 32 | 580 | 6.7 | 76 | 887 | 9.7 |
| 33 | 586 | 6.8 | 77 | 895 | 9.7 |
| 34 | 592 | 6.9 | 78 | 903 | 9.8 |
| 35 | 597 | 6.9 | 79 | 911 | 9.8 |
| 36 | 603 | 7 | 80 | 920 | 9.8 |
| 37 | 609 | 7.1 | 81 | 928 | 9.8 |
| 38 | 615 | 7.2 | 82 | 936 | 9.8 |
| 39 | 621 | 7.3 | 83 | 944 | 9.8 |
| 40 | 627 | 7.3 | 84 | 952 | 9.8 |

| 41 | 633 | 7.4 | 85 | 961 | 9.9 |
| 42 | 639 | 7.5 | 86 | 969 | 9.9 |
| 43 | 645 | 7.6 | 87 | 977 | 9.9 |
| 44 | 652 | 7.7 | 88 | 985 | 9.9 |

# Appendix B: NMF update rules convergence proof

The multiplication update rule in the NMF algorithm has been proved to be nonincreasing in (Lee and Seung 2001). The mathematical reasoning is provided here.

Define $G(h, h')$ as an auxiliary function for $F(h)$, if the conditions:

$$G(h, h') \geq F(h), \quad G(h, h) = F(h) \tag{B-1}$$

are satisfied. If $G$ is such an auxiliary function, then F is nonincreasing under the update below:

$$h^{t+1} = minimize\ G(h, h'). \tag{B-2}$$

Because:

$$F(h^{t+1}) \leq G(h^{t+1}, h^t) \leq G(h^t, h^t) = F(h^t). \tag{B-3}$$

Only if $h^t$ is a local minimum of $G(h^t, h^t)$, $F(h^{t+1}) = F(h^t)$. If the derivatives of $F$ exist and are continuous in a small neighborhood of $h^t$, the derivatives $\nabla F(h^t) = 0$. Iterating the update in Equation (B-2) derives a sequence of estimates that converge to a local minimum $h_{min} = \arg \min F(h)$ of the objective function. Figure B-1 provides an illustration of the converging.
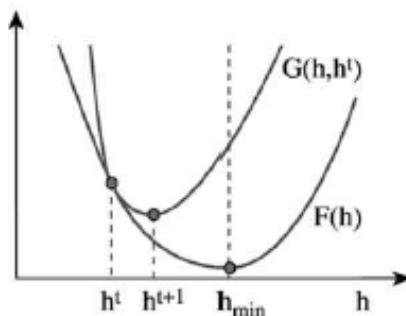


Figure B-1: Converging illustration.

Then, finding an appropriate auxiliary function $G(h, h')$ for Equation (2-6), the multiplicative update rule in Equation (2-7) can be justified.

If $K(h^t)$ is a diagonal matrix:

$$K_{ab}(h^t) = \delta_{ab}(W^T W h^t)_a / h_a^t. \tag{B-4}$$

then, we have

$$G(h, h^t) = F(h^t) + (h - h^t)^T \nabla F(h^t) + \frac{1}{2}(h - h^t)^T K(h^t)(h - h^t). \tag{B-5}$$

which is an auxiliary function of the objective function below:

$$F(h) = \frac{1}{2} \sum_i (v_i - \sum_a W_{ia} h_a)^2. \tag{B-6}$$

It is obvious that $G(h, h) = F(h)$. To prove $G(h, h') \geq F(h)$, we first use Equation (B-5) to minus the formula below:

$$F(h) = F(h^t) + (h - h^t)^T \nabla F(h^t) + \frac{1}{2}(h - h^t)^T (W^T W)(h - h^t). \tag{B-7}$$

Then we have the formula below:

$$(h - h^t)^T (K(h^t) - W^T W)(h - h^t) \tag{B-8}$$

Consider the matrix: $M_{ab}(h^t) = h_a^t(K(h^t) - W^T W)_{ab} h_b^t$, which is a rescaling of Equation (B-8). The $K - W^T W$ is positive semidefinite if and only if $M$ satisfies:

$$v^T M v = \sum_{ab} v_a M_{ab} v_b \tag{B-9}$$

$$= \sum_{ab} h_a^t (W^T W)_{ab} h_b^t v_a^2 - v_a h_a^t (W^T W)_{ab} h_b^t v_b \tag{B-10}$$

$$= \sum_{ab} (W^T W)_{ab} h_a^t [\frac{1}{2}v_a^2 + \frac{1}{2}v_b^2 - v_a v_b] \tag{B-11}$$

$$= \frac{1}{2}\sum_{ab} (W^T W)_{ab} h_a^t (v_a - v_b)^2 \tag{B-12}$$

$$\geq 0 \tag{B-13}$$

The minimum of $G(h, h^t)$ with respect to $h$ is determined by setting the gradient to zeros:

$$\frac{dG(h, h^t)}{dh_a} = -\sum_i v_i \frac{W_{ia} h_a^t}{\sum_b W_{ib} h_b^t} \frac{1}{h_a} + \sum_i W_{ia} = 0 \tag{B-14}$$

Thus, the update rule of Equation (B-2) is:

$$h_a^{t+1} = \frac{h_a^t}{\sum_b W_{kb}} \sum_i \frac{v_i}{\sum_b W_{ib} h_b^t} W_{ia} \tag{B-15}$$

Define:

$$G(h, h^t) = \sum_i (v_i log v_i - v_i) + \sum_{ia} W_{ia} h_a$$

$$- \sum_{ia} v_i \frac{W_{ia} h_a^t}{\sum_b W_{ib} h_b^t} (log W_{ia} h_a - log \frac{W_{ia} h_a^t}{\sum_b W_{ib} h_b^t}) \tag{B-16}$$

This is an auxiliary function for

$$F(h) = \sum_i v_i log \frac{v_i}{\sum_a W_{ia} h_a^t} - v_i + \sum_a W_{ia} h_a. \tag{B-17}$$

It is obvious that $G(h, h) = F(h)$. To prove $G(h, h') \geq F(h)$, we use the convexity of the log function to derive the inequality:

$$- \log \sum_a W_{ia} h_a \leq - \sum_a \alpha_a \log \frac{W_{ia} h_a}{\alpha_a}. \qquad \text{(B-18)}$$

where $\alpha_a = \frac{W_{ia} h_a^t}{\sum_b W_{ib} h_b^t}$. Equation (B-18) holds for all nonnegative $\alpha_a$ that sum to unity.

Then from the inequality it follows that $G(h, h') \geq F(h)$.

# References

Lee, D. D. & H. S. Seung. 2001. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, 556-562.

# Appendix C: Theoretical justification of the LPP and NPE algorithms

In the previous section 4.2.1.1 LPP and NPE, both the locality preserving projection (LPP) and neighborhood preserving embedding (NPE) algorithms convert the minimization problem into eigenproblem. Here, the theoretical justification for this conversion is provided. See (Chung 1997) for a comprehensive reference.

Recall the objective of a LPP algorithm. For two data points that are close to each other in the original space, the LPP algorithm aims to keep them close in the projected feature space. To ensure this objective, the minimization problem as below is used:

$$\min \Sigma_{ij}(y_i - y_j)^2 W_{ij}, \tag{C-1}$$

where $y_i$ and $y_j$ are data points in the projected feature space; and the $W_{ij}$ is the weight between point $i$ and $j$. The minimization function forces a heavy penalty if neighboring points $x_i$ and $x_j$ are mapped far apart in the projected feature space. Thus, by minimizing the function, the neighboring points stay close in the projected feature space.

Given a transformation vector $\mathbf{a}$ that has $y^T = \mathbf{a}^T X$. The $i$th column vector of $X$ is $x_i$. The minimization function (C-1) can be reformulated as below:

$$\frac{1}{2}\sum_{ij}(y_i - y_j)^2 W_{ij} = \frac{1}{2}\sum_{ij}(y_i^2 + y_j^2 - 2y_i y_j)W_{ij} \tag{C-2}$$

$$= \frac{1}{2}(\sum_i y_i^2 D_{ii} + \sum_j y_j^2 D_{jj} - 2\sum_{i,j} y_i y_j W_{ij})$$

$$= \frac{1}{2}(2\sum_i y_i^2 D_{ii} - 2\sum_{i,j} y_i y_j W_{ij}) = \sum_i \mathbf{a}^T x_i D_{ii} x_i^T \mathbf{a} - \sum_{ij} \mathbf{a}^T x_i W_{ij} x_j^T \mathbf{a}$$

$$= \mathbf{a}^T X(D - W)X^T \mathbf{a} = \mathbf{a}^T XLX^T \mathbf{a},$$

where the Laplacian matrix $L = D - W$; and $D_{ii} = \sum_j W_{ji}$ is a diagonal matrix with column sums of W. Matrix $D$ is a natural measure on the data points. The bigger the value $D_{ii}$ is, the more impact from $y_i$ is put on the minimization function. In order to remove an arbitrary scaling factor in the embedding, an additional constraint is imposed as follows:

$$y^T D y = 1 \Rightarrow \boldsymbol{a}^T X L X^T \boldsymbol{a} = \boldsymbol{1}, \tag{C-3}$$

Then, the minimization problem reduces to the function below:

$$\min \boldsymbol{a}^T X L X^T \boldsymbol{a}, \text{ s.t. } \boldsymbol{a}^T X L X^T \boldsymbol{a} = \boldsymbol{1}. \tag{C-4}$$

The transformation vector **a** that minimizes the objective function is given by the minimum eigenvalue solution to the generalized eigenvalue problem:

$$X L X^T \boldsymbol{a} = \lambda X D X^T \boldsymbol{a}. \tag{C-5}$$

In the case of the NPE algorithm, the objective function is to keep the reconstruction error among each neighborhood small:

$$\min \sum_{ij} (y_i - \sum_j W_{ij} y_j)^2. \tag{C-6}$$

Suppose a transformation vector **a** that has $y^T = \boldsymbol{a}^T X$. The $i$th column vector of $X$ is $x_i$. Define

$$z_i = y_i - \sum_j W_{ij} y_j. \tag{C-7}$$

Then, we have the vector form:

$$\boldsymbol{z} = \boldsymbol{y} - \boldsymbol{W} \boldsymbol{y} = (\boldsymbol{I} - \boldsymbol{W}) \boldsymbol{y}. \tag{C-8}$$

where $I = diag(1, \dots, 1)$.

The NPE objective function can be reformulated as below:

$$\sum_i (y_i - \sum_j W_{ij}y_j)^2 = \sum_i (z_i)^2 \tag{C-9}$$

$$= \boldsymbol{z}^T\boldsymbol{z} = y^T(I - W)^T(I - W)y$$

$$= \boldsymbol{a}^T X(I - W)^T(I - W)X^T \boldsymbol{a}$$

$$= \boldsymbol{a}^T XMX^T \boldsymbol{a}.$$

where $M = (I - W)^T(I - W)$. In order to remove an arbitrary scaling factor in the embedding, an additional constraint is imposed as follows:

$$y^T y = 1 \Rightarrow \boldsymbol{a}^T XX^T \boldsymbol{a} = \boldsymbol{1}, \tag{C-10}$$

Then, the minimization problem reduces to the function below:

$$\min \boldsymbol{a}^T XMX^T \boldsymbol{a}, \text{ s.t. } \boldsymbol{a}^T XX^T \boldsymbol{a} = \boldsymbol{1}. \tag{C-11}$$

The transformation vector **a** that minimizes the objective function is given by the minimum eigenvalue solution to the generalized eigenvalue problem:

$$XMX^T \boldsymbol{a} = \lambda XMX^T \boldsymbol{a}. \tag{C-12}$$

### References

Chung, F. R. K. 1997. *Spectral graph theory*. American Mathematical Soc.

# Appendix D: Matlab code

```matlab
function [W, H, opt] = NMF_sparse_volume_robust(X, K, opt)

% Non-Negative Matrix Factorisation with three different constraints
%
% Linear model:     X = W*H + e,     s.t. X>=0, W>=0, H>=0
%
% Input:
% - X (M,N) : M (dimensionality) x N (samples) non negative input
matrix
% - K       : Number of endmembers
% - opt
%   .lambda : Weight for sparsity on H
%   .alpha  : Weight for min. volume constraint on W
%   .omega  : Weight for nonlinearity
%   .W.init : Initialization of W (optional)
%   .W.step : Initial stepsize for W (optional)
%   .H.init : Initialization of H (optional)
%   .H.step : Initial stepsize for H (optional)
%   .R.init : Initialization of R (optional)
%   .R.step : Initial stepsize for R (optional)
%   .tol    : Convergence criteria tolerance
%   .maxIt  : Maximum number of iterations to run
%
% Output:
% - W       : Vertices/Endmembers,   M x K matrix
% - H       : Fractional abundances, K x N matrix
%
% Copyright (c) 2018: Boyu Feng, July 2018.
% Based on 2009: Morten Arngren, September 2009.
%

tic
try opt.lambda;     catch opt.lambda    = 0;     end % Sparsity for H.
try opt.alpha;      catch opt.alpha     = 0;     end  % Min vol for W.
try opt.omega;      catch opt.omega     = 0;     end % Nonlinearity.
try opt.W.init;     catch opt.W.init    = [];    end % Init. of W.
try opt.W.step;     catch opt.W.step    = 1e-3;  end
try opt.H.init;     catch opt.H.init    = [];    end % Init. of H.
try opt.H.step;     catch opt.H.step    = 1e-3;  end
try opt.R.init;     catch opt.R.init    = 0;     end % Init. of H.
try opt.maxIt;      catch opt.maxIt     = 1000;  end
try opt.maxItLocal; catch opt.maxItLocal = 100;  end


% Init W or H from outside
if ~isempty(opt.W.init)
  W(:,1:size(opt.W.init,2)) = opt.W.init;
else
  W = X(:,ceil(N*rand(K,1)));
end
if opt.H.init
```

```matlab
  H = opt.H.init;
else
  H = rand(K,N);
end
if opt.R.init
  R = opt.R.init;
else
  R = rand(D,N);
end

Rscale  = sum(sum(W*H));
sqrnorm = sqrt(Rscale/Xscale);
H       = H/sqrnorm;
W       = W/sqrnorm;
R       = R/sqrnorm;
H       = H ./ repmat(sum(H),K,1);   % Sum-to-one normalization

XX2     = sum(sum(X.^2));
meanX   = mean(X,2);
meanX = repmat(mean(X,2), 1, K);

% initial min vol constraint term
J   = det(W'*W);

% Calculate initial error
errOld = XX2 - 2*sum(sum(W'*X.*H)) + sum(sum( W'*W.*(H*H') ));
errOld = 0.5*(errOld + opt.lambda*sum(sum(H)) + opt.alpha*J +
opt.omega*sum(sqrt(sum(R.^2,1))));
Err     = errOld;

%% Iterate
n = 1; deltaErr = inf; Outerloop = 1;
while Outerloop

  % update R, the outlier term
  R = R.*((X.*(W*H).^(-1))./((W*H).^0 +
opt.omega*R./repmat(sqrt(sum(R.^2,1)),size(X,1),1)));

  %%%%% Update H
  WtX = W'*X;
  WtW = W'*W;

  grad = WtW*H - WtX + opt.lambda;
  grad = grad - repmat(sum(H.*grad),K,1);

  loop = 1; itLocal = 0;
  while loop
    H_        = H - opt.H.step*grad;
    H_(H_<0) = 0;          % Project negative elements to positive
    H_        = H_ ./ repmat(sum(H_),K,1);
    err      = XX2 - 2*sum(sum(WtX.*H_)) + sum(sum( WtW.*(H_*H_') ));
    err      = 0.5*(err + opt.lambda*sum(sum(H)) + opt.alpha *J +
opt.omega*sum(sqrt(sum(R.^2,1))));
```

```matlab
    if err < errOld
      opt.H.step = 1.2*opt.H.step;
      H = H_; loop = 0;
    else
      opt.H.step = 0.5*opt.H.step;
    end
    itLocal = itLocal + 1;
    if itLocal > opt.maxItLocal, loop = 0; end
  end

  % Save current error
  errOld = err;

  %%%%% Update W
  HHt  = H*H';
  XHt  = X*H';
  grad = W*HHt - XHt;
  loop = 1; itLocal = 0;

  WtW  = W'*W;
  J    = det(WtW);
  grad = grad + (tau*2*J*W)/WtW;

    while loop
      W_ = W - opt.W.step*grad;
      if opt.W.nn
        W_(W_<0) = 0;    % Project negative elements to positive
      end
      WtW = W_'*W_;

      % Calc. new regulation term for error estimation.
      J   = det(WtW);

      % Calc. error of new step
      err = XX2 - 2*sum(sum(W_.*XHt)) + sum(sum( WtW.*HHt ));
      err = 0.5*(err + opt.lambda*sum(sum(H)) + opt.alpha*J +
opt.omega*sum(sqrt(sum(R.^2,1))));

      if err < errOld
        opt.W.step = 1.2*opt.W.step;
        W = W_; loop = 0;
      else
        opt.W.step = 0.5*opt.W.step;
      end
      itLocal = itLocal + 1;
      if itLocal > opt.maxItLocal, loop = 0; end
    end
  end

  % Cals. errors
  deltaErr = Err-err;
  Err      = err;
  errOld   = err;
```

```matlab
    n        = n + 1;
    regWAll  = [regWAll J];

    % Is error low enough to stop?
    Outerloop = abs(deltaErr) > opt.tol & n<opt.maxIt;

end
opt.it = n;

toc(tic)
```

```matlab
function [t]=mvntf_constrained(X,R,L,A,B,C,N,options)
% x: the HSI in XJK where the I J K denote the width height and band
number
% R: the number of endmembers
% L: Rank of each abudance map
% A:initialization of A (I*RL)
% B: initial value of B (J*RL)
% C: initial value of C (K*R)
% N: initial value of Nonlinear (I*JK)
% options:
%     derta:  the sum to one penalty
%     alpha:  the volume penalty
%     lambda: weight for sparsity on abundance
%     omega:  weight for outlinear term
%
% maxiters=options.maxiters;%number of maximum iteration
%
% Copyright (c) 2018: Boyu Feng, July 2018.
% Based on 2017: Yuntao Qian, March 2017.
%


convergeNum=options.convergeNum;%converge number
derta=options.derta;%sum to one penalty
alpha=options.alpha;
lambda=options.lambda;
omega=options.omega;
[I,J,K]=size(X);       % Size of the problem

% Matrix Unfoldings of X
X1=tens2mat(X,1);   % X1 is IxJK
X2=tens2mat(X,2);   % X2 is JxIK (or smaller dimensions if compression
was done)
X3=tens2mat(X,3);   % X3 is KxJI (idem)
Ps=kron(eye(R),ones(L,1)); % Pattern matrix
% LOOP for alternating updates
volConstrain=C-mean(C,2)*ones(1,R);
volConstrain=sum(sum(volConstrain.^2));
btdRes{1}=A;
btdRes{2}=B;
btdRes{3}=C;
abundanceRes=normAbundance(btdRes,R,L);
times=0;
sad=0;
rmse=0;
sumToOne=ones(I,J);
iter=1;
objNew=norm(X1-A*myKr(C,B,ones(1,R),ones(1,R)*L)'-
N,'fro')+0.5*alpha*volConstrain+0.5*lambda*norm(abundanceRes,2);
objhistory=[];
objhistory=[objhistory objNew];
while 1
    oldEndmember=C;
    oldAbundance=abundanceRes;
    objOld=objNew;
```

```matlab
        oldSparsity=sparsity;
        oldSad=sad;
        oldRmse=rmse;
        M=myKr(C,B,ones(1,R),ones(1,R)*L);
        A=A.*((X1*M+derta*sumToOne*B)./(A*M'*M+1e-4+derta*A*B'*B));
        M=myKr(C,A,ones(1,R),ones(1,R)*L);
        B=B.*((X2*M+derta*sumToOne'*A)./(B*M'*M+1e-4+derta*B*A'*A));
        M=kr(B,A)*Ps;
        C=C.*((X3*M)./(C*M'*M));
        volConstrain=C-mean(C,2)*ones(1,R);
        volConstrain=sum(sum(volConstrain.^2));
        if omega==0
            X1_ap=A*myKr(C,B,ones(1,R),ones(1,R)*L)';
        else
            X1_ap=A*myKr(C,B,ones(1,R),ones(1,R)*L)'+N;
        end

        if ~omega==0
            N=N.*((X1.*X1_ap.^0)./(X1_ap.^1 +
omega*N./repmat(sqrt(sum(N.^2,1))+eps,I*J,1)));
        end
        btdRes{1}=A;
        btdRes{2}=B;
        btdRes{3}=C;
        %btdRes{4}=N;
        abundanceRes=normAbundance(btdRes,R,L);
        objNew=0.5*norm(X1-A*myKr(C,B,ones(1,R),ones(1,R)*L)'-
omega*N,'fro')+0.5*derta*norm(sumToOne-
A*B','fro')+0.5*alpha*volConstrain+0.5*lambda*norm(abundanceRes,2);
        objhistory=[objhistory objNew];
        toltemp = abs(objOld - objNew)/objOld;
        fprintf('iter [%d]: obj [%d],sad [%d],rmse [%d], C[%d]\n ',
iter,objNew,sad,rmse,sparsity,norm(C(:),2));
        [sad,allSadDistance,sor]=cosDistance(C,oldEndmember);
        [rmse]=HyperRmse(oldAbundance,abundanceRes,sor);
        if (abs(oldSparsity-sparsity)/oldSparsity<2e-4&&abs(oldSad-
sad)/oldSad<(2e-4)&&abs(oldRmse-rmse)/oldRmse<(2e-4))
            times = times + 1;
        else
            times=0;
        end

        iter = iter+1;
        if times==convergeNum
            t{1}=C;
            t{2}=abundanceRes;
            t{3}=objhistory;
            t{4}=btdRes;
            break;
        end
    end

    %***************************************************************
    *********
```

```matlab
function [U1,U2,U3,S,S1,S2,S3] = mlsvd3(X,size_core)
%MLSVD3 Multilinear singular value decomposition of a third-order
tensor.
[I1,I2,I3]=size(X);
[U1,S1,temp]=svd(reshape(X,I1,I3*I2),'econ'); S1=diag(S1);
[U2,S2,temp]=svd(reshape(permute(X,[2 3 1]),I2,I1*I3),'econ');
S2=diag(S2);
[U3,S3,temp]=svd(reshape(permute(X,[3 1 2]),I3,I2*I1),'econ');
S3=diag(S3);
if nargin==2
    U1=U1(:,1:min(size_core(1),I2*I3));
    U2=U2(:,1:min(size_core(2),I1*I3));
    U3=U3(:,1:min(size_core(3),I1*I2));
end
S=tmprod(tmprod(tmprod(X,U1',1),U2',2),U3',3);
end


%********************************************************************
*********
function X_out = tmprod(X,U,mode)
%TMPROD mode-n tensor-matrix product.
[I,J,K]=size(X);
[M,N]=size(U);
if (mode~=1) && (mode~=2) && (mode~=3)
    error('The input variable mode should be 1, 2 or 3')
end
if N~=size(X,mode)
    error(['The number of columns of the input matrix should be equal
to dimension ',int2str(mode),' of the input tensor'])
end
if mode==1
    X_out = reshape(U*reshape(X,I,J*K) ,M,J,K);
elseif mode==2
    X_out = permute(reshape (U*reshape(permute(X,[2 1 3]),J,I*K),
M,I,K),[2 1 3]);
elseif mode==3
    X_out = permute(reshape (U*reshape(permute(X,[3 1 2]),K,I*J),
M,I,J),[2 3 1]);
end
end


function C = kr(A,B)
%KR Khatri-Rao product.
[I R1]=size(A); J=size(B,1);
C=zeros(I*J,R1);
for j=1:R1
    C(:,j)=reshape(B(:,j)*A(:,j).',I*J,1);
end
end


%********************************************************************
*********
function Mat = kr_part(B,C,partB,partC)
%KR_PART Partition-Wise Kronecker product
[J M]=size(B);
```

```matlab
[K N]=size(C);
if (sum(partB)~=M)
    error(['Error: a matrix with ',int2str(M),' columns can not be
partitioned in such a way'])
end
if (sum(partC)~=N)
    error(['Error: a matrix with ',int2str(N),' columns can not be
partitioned in such a way'])
end
if length(partB)~=length(partC)
    error('Error: the 2 input matrices do not have the same number of
blocks')
end

indB=[0 cumsum(partB)];
indC=[0 cumsum(partC)];
indMat=[0 cumsum(partB.*partC)];

Mat=zeros(J*K,sum(partB.*partC));
for i=1:length(partC)
    Mat(:,indMat(i)+1:indMat(i+1))=fast_kron( B(:,indB(i)+1:indB(i+1))  ,
C(:,indC(i)+1:indC(i+1)));
end
end
```

```matlab
function [U, V, B, eigvalue_U,eigvalue_V, eigvalue_B,posIdx,Y] =
TensorLPP (X, W, options,bLPP)
% TensorLPP: Tensor Locality Preserving Projections
%
% [U, V, eigvalue_U, eigvalue_V, posIdx, Y] = TensorLPP(X, W, options)
%
%              Input:
%                X      - 3D data matrix. X(:,:,i) is the i-th data
%                          sample.
%                W      - Weight matrix.
%                options - Struct value in Matlab. The fields in options
%                          that can be set:
%                             nRepeat    -   The repeat times of the
%                                            iterative procedure.
%                                            Default: 10
%                bLPP    - 0/1 value. 1-TLPP and 0-TNPE
%
%              Output:
%                U, V, B    - Embedding functions, for a new data point
%                             (matrix) x, y = x *1 U *2 V *3 B
%                eigvalue_U - corresponding eigenvalue.
%                eigvalue_V - corresponding eigenvalue.
%                eigvalue_B - corresponding eigenvalue.
%
%                Y          - The embedding results, Each row vector is a
%                             data point. The features in Y has been
%                             sorted that Y(:,i) will be important to
%                             Y(:,j)with respect to the objective
%                             function if i<j
%
%                posIdx     - Resort idx. For a new data sample (matrix)
%
% Copyright (c) 2018: Boyu Feng, July 2018.
% Based on 2007: Deng Cai.
%

if (~exist('options','var'))
   options = [];
else
   if ~strcmpi(class(options),'struct')
       error('parameter error!');
   end
end

if ~isfield(options,'nRepeat')
    options.nRepeat = 10;
else
    options.nRepeat = options.nRepeat; %
end

 [window1,window2,nBand,nSmp] = size(X);

if bLPP
    D = sparse(1:nSmp,1:nSmp,sum(W,2),nSmp,nSmp);
    if nargout == 8
```

```matlab
        [U, V, B, eigvalue_U, eigvalue_V,
eigvalue_B,posIdx,Y]=TensorLGE_3ways_v3(X, W, D, options);
    else
        [U, V, B, eigvalue_U, eigvalue_V,
eigvalue_B,posIdx]=TensorLGE_3ways_v3(X, W, D, options);
    end
else
    D=[];
    if nargout == 8
        [U, V, B, eigvalue_U, eigvalue_V,
eigvalue_B,posIdx,Y]=TensorLGE (X, W, D, options);
    else
        [U, V, B, eigvalue_U, eigvalue_V, eigvalue_B,posIdx]=TensorLGE
(X, W, D, options);
    end
end
```

```matlab
function [U, V, B, eigvalue_U, eigvalue_V, eigvalue_B,posIdx,Y] =
TensorLGE (X, W, D, options)
% TensorLGE: Tensor-based Linear Graph Embedding
%
%            Input:
%              X        -  3D data matrix. X(:,:,i) is the i-th data
%                          sample.
%              W        - Weight matrix.
%              D        - Graph matrix.
%
%              options - Struct value in Matlab. The fields in options
%                        that can be set:
%
%                            nRepeat     -   The repeat times of the
%                                            iterative procedure.
%                                            Default: 10
%
%            Output:
%              %            Output:
%              U, V, B    - Embedding functions, for a new data point
%                           (matrix) x, y = x *1 U *2 V *3 B
%              eigvalue_U - corresponding eigenvalue.
%              eigvalue_V - corresponding eigenvalue.
%              eigvalue_B - corresponding eigenvalue.
%
%              Y          - The embedding results, Each row vector is a
%                           data point. The features in Y has been
%                           sorted that Y(:,i) will be important to
%                           Y(:,j)with respect to the objective
%                           function if i<j
%
%              posIdx     - Resort idx. For a new data sample (matrix) %
%
% Copyright (c) 2018: Boyu Feng, July 2018.
% Based on 2007: Deng Cai.
%

if (~exist('options','var'))
   options = [];
end

if isfield(options,'nRepeat')
    nRepeat = options.nRepeat; %
else
    nRepeat = 10;
end

bD = 1;
if ~exist('D','var') | isempty(D)
    bD = 0;
end

[window1,window2,nBand,nSmp] = size(X);
```

```matlab
if size(W,1) ~= nSmp
    error('W and X mismatch!');
end
if bD & (size(D,1) ~= nSmp)
    error('D and X mismatch!');
end


[i_idx,j_idx,v_idx] = find(W);
if bD
    [Di_idx,Dj_idx,Dv_idx] = find(D);
end


U = eye(window1);
V = eye(window2);
B = eye(nBand);

for repeat = 1:nRepeat
    XVB = zeros(window1,window2*nBand,nSmp);
    for i = 1:nSmp
        a = nmodeproduct(X(:,:,:,i),V',2);
        a = nmodeproduct(a,B',3);
        XVB(:,:,i) = tensor_unfolding(a,1);
    end

    S_vb = zeros(window1,window1);
    D_vb = zeros(window1,window1);

    if bD
        for idx=1:length(Di_idx)
            D_vb = D_vb +
Dv_idx(idx)*XVB(:,:,Di_idx(idx))*XVB(:,:,Dj_idx(idx))';
        end
    else
        for i=1:nSmp
            D_vb = D_vb + XVB(:,:,i)*XVB(:,:,i)';
        end
    end

    if bD
        for idx=1:length(i_idx)
            S_vb = S_vb + v_idx(idx)*(XVB(:,:,i_idx(idx))-
XVB(:,:,j_idx(idx)))*(XVB(:,:,i_idx(idx))-XVB(:,:,j_idx(idx)))';
        end
    else

        for idx=1:length(i_idx)
            S_vb = S_vb + (XVB(:,:,i_idx(idx))-
v_idx(idx)*XVB(:,:,j_idx(idx)))*(XVB(:,:,i_idx(idx))-
v_idx(idx)*XVB(:,:,j_idx(idx)))';
        end

    end

    D_vb = max(D_vb,D_vb');
```

177

```matlab
    S_vb = max(S_vb,S_vb');

    [U, eigvalue_U] = eig(S_vb,D_vb);
    eigvalue_U = diag(eigvalue_U);
    [junk, index] = sort(-eigvalue_U);
    U = U(:, index);
    eigvalue_U = eigvalue_U(index);

    for i = 1:size(U,2)
        U(:,i) = U(:,i)./norm(U(:,i));
    end

    XUB = zeros(window2,window1*nBand,nSmp);
    for i = 1:nSmp
        a = nmodeproduct(X(:,:,:,i),U',1);
        a = nmodeproduct(a,B',3);
        XUB(:,:,i) = tensor_unfolding(a,2);
    end

    S_ub = zeros(window2,window2);
    D_ub = zeros(window2,window2);

    if bD
        for idx=1:length(Di_idx)
            D_ub = D_ub +
Di_idx(idx)*XUB(:,:,Di_idx(idx))*XUB(:,:,Dj_idx(idx))';
        end
    else
        for i=1:nSmp
            D_ub = D_ub + XUB(:,:,i)*XUB(:,:,i)';
        end
    end

    if bD
        for idx=1:length(i_idx)
            S_ub = S_ub + v_idx(idx)*(XUB(:,:,i_idx(idx))-
XUB(:,:,j_idx(idx)))*(XUB(:,:,i_idx(idx))-XUB(:,:,j_idx(idx)))';
        end
    else

        for idx=1:length(i_idx)
            S_ub = S_ub + (XUB(:,:,i_idx(idx))-
v_idx(idx)*XUB(:,:,j_idx(idx)))*(XUB(:,:,i_idx(idx))-
v_idx(idx)*XUB(:,:,j_idx(idx)))';
        end

    end

    D_ub = max(D_ub,D_ub');
    S_ub = max(S_ub,S_ub');

    [V, eigvalue_V] = eig(S_ub,D_ub);
    eigvalue_V = diag(eigvalue_V);
    [junk, index] = sort(-eigvalue_V);
```

```matlab
        V = V(:, index);
        eigvalue_V = eigvalue_V(index);

        for i = 1:size(V,2)
            V(:,i) = V(:,i)./norm(V(:,i));
        end

        XUV = zeros(nBand,window1*window2,nSmp);
        for i = 1:nSmp
            a = nmodeproduct(X(:,:,:,i),U',1);
            a = nmodeproduct(a,V',2);
            XUV(:,:,i) = tensor_unfolding(a,3);
        end

        S_uv = zeros(nBand,nBand);
        D_uv = zeros(nBand,nBand);

        if bD
            for idx=1:length(Di_idx)
                D_uv = D_uv +
Di_idx(idx)*XUV(:,:,Di_idx(idx))*XUV(:,:,Dj_idx(idx))';
            end
        else
            for i=1:nSmp
                D_uv = D_uv + XUV(:,:,i)*XUV(:,:,i)';
            end
        end

        if bD
            for idx=1:length(i_idx)
                S_uv = S_uv+ v_idx(idx)*(XUV(:,:,i_idx(idx))-
XUV(:,:,j_idx(idx)))*(XUV(:,:,i_idx(idx))-XUV(:,:,j_idx(idx)))';
            end
        else

            for idx=1:length(i_idx)
                S_uv = S_uv + (XUV(:,:,i_idx(idx))-
v_idx(idx)*XUV(:,:,j_idx(idx)))*(XUV(:,:,i_idx(idx))-
v_idx(idx)*XUV(:,:,j_idx(idx)))';
            end

        end

        D_uv = max(D_uv,D_uv');
        S_uv = max(S_uv,S_uv');

        [B, eigvalue_B] = eig(S_uv,D_uv);
        eigvalue_B = diag(eigvalue_B);
        [junk, index] = sort(-eigvalue_B);
        B = B(:, index);
        eigvalue_B = eigvalue_B(index);

        for i = 1:size(B,2)
```

```matlab
        B(:,i) = B(:,i)./norm(B(:,i));
    end

end

Y = zeros(window1,window2,nBand,nSmp);
for i = 1:nSmp
    y = nmodeproduct(X(:,:,:,i),U',1);
    y = nmodeproduct(y,V',2);
    Y(:,:,:,i) = nmodeproduct(y,B',3);
end

Y = tensor_unfolding(Y,4);

if bD
    DPrime = sum((Y'*D)'.*Y,1);
else
    DPrime = sum(Y.*Y,1);
end
LPrime = sum((Y'*W)'.*Y,1);

DPrime(find(DPrime < 1e-14)) = 10000;
LaplacianScore = LPrime./DPrime;

[dump,posIdx] = sort(-LaplacianScore);

if nargout == 8
    Y = Y(:,posIdx);
end
```

```matlab
function [WRCM_matrix] = weighted_region_covariance_unite(I,l)

%  WRCM_matrix(x,y,:,:) : weighted covariance matrix at position
(x+l,y+l).
%  l: half the side length of square in which the covariance
%     region descriptor is computed, i.e. region is square with
%     side length (2l+1)
%
% Copyright (c) 2018: Boyu Feng, July 2018.
%

[m, n, d] = size(I);

I2 = zeros(m+2*l,n+2*l,d);
I2(l+1:m+l,1+l:n+l,:)=I;

for i = 1:l
    I2(i,1+l:n+l,:) = I2(l+1-i+l,1+l:n+l,:);
end
for i = m+l:m+2*l
    I2(i,1+l:n+l,:) = I2(m+l-(i-(m+l)),1+l:n+l,:);
end

for i = 1:l
    I2(:,i,:) = I2(:,l+1-i+l,:);
end
for i = n+l:n+2*l
    I2(:,i,:) = I2(:,n+l-(i-(n+l)),:);
end

[m, n, d] = size(I2);

WRCM_matrix = zeros(m-2*l,n-2*l,d,d);

options = [];

for i = 1+l:m-l
    for j = 1+l:n-l
        WRCM_matrix(i-l,j-l,:,:)=weighted_region_covariance_block(I2(i-
l:i+l,j-l:j+l,:),options,l);
    end
end
WRCM_matrix = reshape(WRCM_matrix,(m-2*l)*(n-2*l),d,d)*10;
```

```matlab
function [WRCM] = weighted_region_covariance_block(featureBlock,options)
% Region covariance matrix (GCM) algorithm
%
% Inputs:
%       featureBlock        band*m*n
%       options             options
%
% Copyright (c) 2018: Boyu Feng, July 2018.
% Based on 2008: Yanwei Pang, July 2018.
%

    [m,n,d] = size(featureBlock);
    featureBlock = reshape(featureBlock,m*n,d);
    mu = mean(featureBlock,1);
    cPixel = featureBlock(ceil(m*n/2),:);
    symm = @(X) .5*(X+X');

    W = EuDist2(featureBlock);
    W = W/100000;
    W = exp(-W.^2);

    if ~isfield(options, 'spd_projection')
        spd_projection = true;
    else
        spd_projection = options.spd_projection;
    end


    WRCM = zeros(d,d);
    for i = 1:m*n
        for j = 1:m*n
            WRCM = WRCM + W(i,j)*(featureBlock(i,:)-
featureBlock(j,:))'*(featureBlock(i,:)-featureBlock(j,:));
        end
    end
    WRCM = WRCM/(m*n);
```

# Appendix E: Copyright releases from publications

Chapter 2

Feng, B. & Wang, J. (in press). Evaluation of unmixing methods for impervious surface area extraction from simulated EnMAP imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.*

3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

# Curriculum Vitae

**Name:**          Boyu Feng

**Post-secondary Education and Degrees:**

China University of Geosciences
Wuhan, Hubei, China
2008-2012 B.E.

University of Kansas
Lawrence, Kansas, U.S.A
2012-2014 M.A.

University of Western Ontario
London, Ontario, Canada
2014-2018 Ph.D.

**Honors and Awards:**

Lawry's scholor, China University of Geosciences
2009-2011

Honor Graduates, University of Kansas
2014

**Related Work Experience**

Team Leader
Chinese undergraduate mathematical contest in modeling
2011

Research Assistant
Center for Remote Sensing of Ice Sheet (CReSIS)
2012-2014

Research Assistant
University of Western Ontario
2014-2018

Teaching Assistant
University of Western Ontario
2014-2018

**Publications:**

a) *Articles in Refereed Journals*

Feng, B. & Wang, J. (in preparation). Patch-based and tensor-patch-based dimension reduction methods for hyperspectral images.

Feng, B. & Wang, J. (under 1st round revision). Constrained nonnegative tensor factorization for spectral unmixing of hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters*.

Feng, B. & Wang, J. (in press). Evaluation of unmixing methods for impervious surface area extraction from simulated EnMAP imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.*

Feng, B., Braaten, D., Paden, J., & Gogineni, P. (2016). Firn stratigraphic genesis in early spring: evidence from airborne radar. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*(6), 2429-2435.

Feng, B. (2011) Natural Fires Research Progress of Remote Sensing Monitoring. *Public Communication of Science & Technology*, 14: 228-229 (in Chinese).

## b) *Full Papers in Conference Proceedings*

Feng, B. & Wang, J. Impervious surface area extraction using simulated EnMAP imagery. *Proceedings of the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2017),* Fort Worth, Texas, USA. July 23-28, 2017. 4 p.

## c) *Technical Proposals & Reports*

Application of ArcGIS pro in urban flooding forecast and emergency response (Granted)

Engage with ESRI Canada                                            Feb, 2018

Urban flooding modelling using earth observation data – a case study of cities in Great Lakes Region (P: 1-36)

Canada Centre for Remote Sensing, Centre for Mapping and Earth Observation        Apr, 2017

Surrey landuse/land cover mapping using airborne data report (P 1-3)

Government of City of Surrey, BC                                    Apr, 2016

Earth observation based extraction and application of digital surface model (DSM) over urban areas (P 1:33)

Canada Centre for Remote Sensing, Centre for Mapping and Earth Observation        Mar, 2016

## d) *Conference presentations*

Feng, B., Zeng, C., Wang, J. and Zhang, Y. Urban flash flood risk mapping using hydrodynamic and GIS models. Presented at the GIS in Education and Research Conference, Toronto, Ontario, Canada. October 11, 2017.

Feng, B., Zeng, C., Wang, J. and Zhang, Y. Urban flooding risk mapping using hydraulic and GIS models. Presented at the 25th International Conferences on Geoinformatics, Buffalo, NY, USA. August 2-4, 2017.

Feng, B. and Wang, J. Impervious surface area extraction using simulated EnMAP imagery. Presented at *the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2017),* Fort Worth, Texas, USA. July 23-28, 2017.

Feng, B., Zeng, C., Wang, J. and Zhang, Y. Patterns and changes of urban flash flood risks in two Canadian urban centers. Presented at the Earth Observation Summit 2017, Montreal, Quebec, Canada. June 20-22, 2017.

Feng, B., Huang, Y., and Wang, J. Road racking area percentage evaluation using airborne hyperspectral imagery. Presented at the Western Research Forum, London, Ontario, Canada. March 10, 2017.

Feng, B., Huang, Y., and Wang, J. The influence of endmember selection method in extracting impervious surface from airborne hyperspectral imagery. Presented at the American Geophysical Union's 49th annual fall meeting, San Francisco, California. December 12-16, 2016.

Feng, B., Braaten, D., Gogineni, P., Paden, J., Leuschen, C., and Purdon, K. The use of Snow Radar in West Antarctic ice sheet annual snow accumulation study. Presented at the IEEE International Geoscience and Remote Sensing Symposium, Milan, Italy. July 26-31, 2015.

Feng, B., Braaten, D., Gogineni, P., Paden, J., Leuschen, C., and Purdon, K. Linking the annual variation of Snow Radar-derived accumulation in West Antarctica to long-term Automatic Weather Station measurements. Presented at the American Geophysical Union's 46th annual fall meeting, San Francisco, California, USA. December 9-13, 2013.

Feng, B. and Li, X. Urban inundation python script toolbox for ArcGIS. Presented at the GIS Day at University of Kansas, Lawrence, Kansas, USA. November 20, 2013.