



Norwegian University  
of Life Sciences

**Master's Thesis 2024 30 ECTS**  
Faculty of Science and Technology

# **Phase Segregation Mapping in Perovskite Solar Cells Using Hyperspectral Photoluminescence Imaging**

Håkon Holte

Environmental Physics and Renewable Energy



## Preface and Acknowledgements

This master's thesis concludes my time as a master's student in Environmental Physics and Renewable Energy at the Norwegian University of Life Sciences (NMBU). During the last five years, I have truly enjoyed the opportunity and freedom to delve into a wide variety of exciting topics, such as quantum physics, mathematical modeling, deep learning, image analysis, meteorology, and solar energy.

A master's thesis is more of a team effort rather than an individual achievement. In that regard, I want to thank my supervisor, Espen Olsen, and my co-supervisor, Ingunn Burud, for interesting discussions and for giving me valuable feedback during the project. I also wish to thank Fredrik Arnesen Stulen for giving me helpful tips and feedback on my writing, and Andreas Svarstad Flø for helping me with the hyperspectral imaging setup.

I would like to express my deepest appreciation to my co-supervisor, Ivar Loland Råheim. You have taught me a lot about perovskite solar cells and experimental work, in addition to giving me valuable input on my writing and the project in general. I am really looking forward to seeing the results of your research!

Writing this thesis would have been significantly harder without the support from my friends and family. I am extremely grateful for the support and care from my parents, Hilde and Torolf, and my siblings Erling, Cecilie, Magnus, and Ingrid. I am also grateful to my fellow students here at NMBU and my friends in Studentorkesteret STRYK for motivating and inspiring me.

Finally, I would like to extend my deepest gratitude to my grandmother Momme, who sadly passed away this spring. We looked forward to sharing this moment, but you will forever have a special place in my heart. This one is for you.

*Håkon Holte*

Ås, June 2024



---

## Abstract

Hybrid organic-inorganic lead halide perovskites represent a class of novel solar cell materials that have recently gained interest due to their excellent optoelectronic properties. However, their poor stability and fast degradation hinder commercialization. This work investigates one instability mechanism, light-induced phase segregation, by analyzing hyperspectral photoluminescence (HSPL) images of mixed-halide perovskite solar cells.

The present study aims to use HSPL imaging and accompanying data analysis techniques to map light-induced phase segregation in mixed-halide perovskite solar cells. To achieve this, planar regular (n-i-p) perovskite solar cells were fabricated in REALTEK's laboratories. HSPL images of these cells were then analyzed using data analysis pipelines comprising some preprocessing steps in addition to principal component analysis (PCA) and multivariate curve resolution (MCR). Supplementary X-ray fluorescence (XRF) images of the samples were also taken in an attempt to assess whether the HSPL images can be used for mapping phase segregation-caused regions of higher and lower concentrations of the perovskite's halide species (iodide and bromide).

The samples' photoluminescence (PL) spectra comprise several peaks and shoulders that may originate from phases of different halide stoichiometries. However, it can not be concluded that the observed PL signals are caused by phase segregation since the XRF images from the present XRF setup could not be used for validation. Also, the literature suggests that similar spectral features may be caused by other mechanisms, like re-absorption effects. A thorough analysis of the origins of the various PL signals must be conducted, and a suitable characterization method for validation of the HSPL analysis must be in place before the method can be developed further. Since the literature suggests that the iodide-rich domains are small compared to the resolutions of the present XRF and HSPL setups, increasing the resolution may be necessary to facilitate the localization of such domains.

---

## Sammendrag

Hybride organisk-uorganiske blyhalidperovskitter representerer en ny klasse solcellematerialer som den siste tiden har vekket stor interesse som følge av deres utmerkede optoelektroniske egenskaper. Likevel forhindrer deres dårlige stabilitet og raske degradering kommersialisering. Dette arbeidet undersøker én ustabilitetsmekanisme, lysindusert fasesegregering, ved å analysere hyper-spektrale fotoluminescensbilder (HSPL) av fler-halid perovskittsolceller.

Studien tar sikte på å bruke HSPL-avbildning og tilhørende dataanalysemetoder for å kartlegge lysindusert fasesegregering i fler-halid perovskittsolceller. For å oppnå dette ble plane perovskittsolceller med n-i-p arkitektur produsert i REALTEKs laboratorier. HSPL-bilder av disse cellene ble deretter analysert ved hjelp av en kombinasjon av ulike preprosesseringssteg, prinsipalkomponentanalyse (PCA) og multivariat kurveoppløsning (MCR). Supplementære røntgenfluorescensbilder (XRF) tas også av prøvene i et forsøk på å vurdere om HSPL-bildene kan brukes til å kartlegge områder med høyere og lavere konsentrasjoner av perovskittens halider (jod og brom) forårsaket av fasesegregering.

Prøvenes fotoluminescensspektre (PL) inneholder flere topper og skuldre som kan stamme fra faser med ulike halidstøkiometrier. Likevel kan det ikke konkluderes med at de observerte PL-signalene skyldes fasesegregering, siden XRF-bildene som ble tatt med det aktuelle XRF-oppsettet ikke kunne brukes til validering. Litteraturen foreslår at liknende spektrale fenomener også kan skyldes andre mekanismer, slik som reabsorpsjonseffekter. En grundig analyse av opprinnelsen til de ulike PL-signalene må utføres, i tillegg til at en passende karakteriseringsmetode for å validere HSPL-analysen må på plass før metoden kan utvikles videre. Siden litteraturen foreslår at de jod-rike domeneene er små sammenliknet med oppløsningen til de nåværende XRF- og HSPL-oppsettene, kan en økning av oppløsningen være nødvendig for å kunne lokalisere slike domener.

---

## Use of Artificial Intelligence (AI) Tools

Grammarly, ChatGPT 3.5, and Sikt KI-chat (with the gpt-4-8k model) are the Artificial Intelligence (AI) tools that were utilized in this work. The list below summarizes how these tools were used and gives examples of prompts written to these AI language models.

- Grammarly was used for performing augmented grammar and spell checks.
  - For instance, Grammarly marks misspelled words and incorrect use of punctuation in the text editor.
  - Moreover, it provides suggestions on how to improve the sentence structure.
- ChatGPT 3.5 and Sikt KI-chat were used to find the most appropriate word for use in a sentence.
  - Example prompt: "Should I write X or Y in this context?"
- ChatGPT 3.5 and Sikt KI-chat were used to translate terms and phrases from Norwegian to English and vice versa.
  - Example prompt: "Translate X from Norwegian to English."
  - Note! These tools were not used to translate complete sentences or longer text segments for use in this text (the thesis). They were exclusively used to translate single words or phrases.
  - The translations were always double-checked with a dictionary if something was unclear.
- ChatGPT 3.5 and Sikt KI-chat were used to help generate Python code.
  - Example prompt: "How can a three-dimensional numpy array be transformed into a matrix?"
  - Example prompt: "What does this error message mean?"
  - Example prompt: "How can the font size of the x-axis label in a matplotlib plot be adjusted?"
- ChatGPT 3.5 and Sikt KI-chat were used to generate Latex code.
  - Example prompt: "How can I make a table containing...?"

---

## **A Note on Image Quality in Printed Copies of the Thesis**

Some printers and/or printer settings may reduce the detail level in printed copies of this work's images. Hence, to obtain satisfactory image quality and to preserve important details, the recommendation is to either use appropriate printer settings (a sufficiently high DPI) or to view the PDF digitally.

Moreover, the X-ray fluorescence (XRF) images in Section 5.4 are often difficult to read in printed copies due to their colors. Reading a digital copy makes interpreting the figures easier. Another reason to have a look at the digital copy is that the XRF software produced plots with too small font sizes. This has made reading the axis labels in Figure 5.8 difficult in printed copies.



## List of Abbreviations

*BNM* background noise matrix.

*LoF* lack of fit.

*PLoF* pixel-wise lack of fit.

**AR** alternating regression.

**BB** band-to-band.

**CCD** charge-coupled device.

**CMOS** complementary metal-oxide semiconductor.

**EL** electroluminescence.

**ETL** electron transport layer.

**FA** formamidinium.

**HIT** heterojunction with intrinsic layer.

**HSPL** hyperspectral photoluminescence.

**HTL** hole transport layer.

**IEA** International Energy Agency.

**ITO** indium tin oxide.

**MA** methylammonium.

**MCR** multivariate curve resolution.

**MCR-ALS** multivariate curve resolution-alternating least squares.

**OLS** ordinary least-squares.

**PCA** principal component analysis.

**PL** photoluminescence.

**PSC** perovskite solar cell.

**PV** photovoltaic.

**SEM** scanning electron microscopy.

**SRH** Shockley-Read-Hall.

**TCO** transparent conductive oxide.

## List of Abbreviations

---

**TEM** transmission electron microscopy.

**XRD** X-ray diffraction.

**XRF** X-ray fluorescence.

# Contents

<b>Preface and Acknowledgements</b>	<b>i</b>
<b>Abstract</b>	<b>iii</b>
<b>Sammendrag</b>	<b>iv</b>
<b>Use of Artificial Intelligence (AI) Tools</b>	<b>v</b>
<b>A Note on Image Quality in Printed Copies of the Thesis</b>	<b>vi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Objective and Project Description . . . . .	3
1.2 Research questions . . . . .	4
<b>2 Theory</b>	<b>5</b>
2.1 Solar Cell and Semiconductor Fundamentals . . . . .	5
2.1.1 Direct and indirect bandgap semiconductors . . . . .	6
2.1.2 Recombination mechanisms . . . . .	7
2.2 Perovskites . . . . .	10
2.2.1 Perovskite structure . . . . .	10
2.2.2 Bandgap and optoelectronic properties . . . . .	11
2.2.3 Light-induced phase segregation . . . . .	13
2.3 Perovskite Solar Cells . . . . .	15
2.3.1 Cell architecture . . . . .	15
2.3.2 Working principles of the perovskite solar cells . . . . .	16
2.4 Hyperspectral Photoluminescence (HSPL) Imaging . . . . .	17
2.4.1 Photoluminescence (PL) and electroluminescence (EL) . . . . .	17
2.4.2 Hyperspectral Imaging . . . . .	18
2.4.3 Line scan/push-broom camera . . . . .	19
2.5 X-ray Fluorescence (XRF) Imaging . . . . .	21
2.6 Data Analysis . . . . .	22
2.6.1 Principal Component Analysis (PCA) . . . . .	22
2.6.2 Multivariate Curve Resolution (MCR) . . . . .	23
<b>3 Experimental</b>	<b>26</b>
3.1 Perovskite Solar Cell Fabrication . . . . .	26
3.1.1 Substrate cleaning . . . . .	26
3.1.2 Electron transport layer (ETL) . . . . .	26
3.1.3 Perovskite precursor solution . . . . .	27
3.1.4 Perovskite film deposition . . . . .	28

3.1.5	Hole transporting layer (HTL) . . . . .	28
3.2	Hyperspectral Photoluminescence (HSPL) Imaging . . . . .	29
3.2.1	Instruments . . . . .	29
3.2.2	Imaging setup . . . . .	30
3.2.3	Image acquisition procedures . . . . .	31
3.2.4	Overview of samples and hyperspectral images studied in this work . . . . .	32
3.3	X-ray Fluorescence (XRF) Imaging . . . . .	33
<b>4</b>	<b>Data Processing</b> . . . . .	<b>34</b>
4.1	Software . . . . .	34
4.2	Preprocessing . . . . .	34
4.2.1	Background noise correction . . . . .	35
4.2.2	Jacobian conversion from wavelength (nm) to energy (eV) units . . . . .	36
4.2.3	Scaling the intensity values by normalizing the areas un- derneath the pixel spectra . . . . .	37
4.3	Data Analysis Pipelines . . . . .	40
4.3.1	Pipeline 1: PCA . . . . .	41
4.3.2	Pipeline 2: PCA + MCR . . . . .	41
<b>5</b>	<b>Results and Discussion</b> . . . . .	<b>45</b>
5.1	Estimating the Perovskite Samples' Bandgaps . . . . .	45
5.2	Time Series of Single-Line PL Spectra (Sample 2) . . . . .	46
5.3	Raw Spatial HSPL Image (Sample 1) . . . . .	49
5.3.1	Accumulated PL intensities . . . . .	49
5.3.2	Mean PL spectrum . . . . .	50
5.3.3	Comparing PL intensities from sample 1 and sample 3 . . . . .	51
5.4	XRF Images (Sample 1) . . . . .	53
5.5	Preprocessing (Sample 1) . . . . .	57
5.5.1	Background noise correction . . . . .	57
5.5.2	Pixel spectra normalization . . . . .	58
5.6	Mean PL Spectrum in the Energy Domain (Sample 1) . . . . .	59
5.7	PCA (Sample 1) . . . . .	60
5.7.1	PCA without normalization . . . . .	60
5.7.2	PCA with normalization . . . . .	64
5.7.3	Discussion: comparing PCA with and without normal- ization . . . . .	67
5.7.4	Discussion: spatial information extracted by the princi- pal components . . . . .	67

## CONTENTS

---

5.7.5	Summary of the PCA analysis . . . . .	70
5.8	Exploring and Assessing Different MCR Models (Sample 1) . . . . .	71
5.8.1	Choosing the number of MCR components . . . . .	71
5.8.2	Comparing the candidate MCR models . . . . .	77
5.8.3	Pixel-wise "lack of fit" ( <i>PLoF</i> ) of candidate models . . . . .	79
5.8.4	Comparing the candidate models by decomposing the PL spectra for selected regions . . . . .	80
5.9	Interpreting the MCR results (Sample 1) . . . . .	85
5.9.1	Investigating the pure spectra of the MCR model with four components . . . . .	85
5.9.2	Comparison with the three-component model . . . . .	87
5.9.3	Discussion: segregated-phase stoichiometries reported in the literature . . . . .	87
5.9.4	Discussion: PL signals not directly related to phase seg- regation . . . . .	89
5.9.5	Discussion: correlated PL peaks . . . . .	90
5.10	Pros and Cons Associated with PCA and MCR . . . . .	90
5.10.1	PCA . . . . .	90
5.10.2	MCR . . . . .	91
5.11	Limitations and Uncertainties . . . . .	92
<b>6</b>	<b>Conclusions</b> . . . . .	<b>93</b>
6.1	Outlook and Further Work . . . . .	94
	<b>References</b> . . . . .	<b>101</b>
	<b>Appendices</b> . . . . .	<b>102</b>
<b>A</b>	<b>Python code used for data preprocessing</b> . . . . .	<b>102</b>
A.1	Reading the hyperspectral EL image . . . . .	102
A.2	Reading the hyperspectral PL images (Specim) . . . . .	103
A.2.1	Background noise correction . . . . .	104
A.3	Jacobian conversion from wavelength (nm) to energy (eV) units . . . . .	104
A.4	Scaling the intensity values by normalizing the areas underneath the pixel spectra . . . . .	105
<b>B</b>	<b>Python code used for data analysis</b> . . . . .	<b>107</b>
B.1	PCA . . . . .	107
B.2	MCR . . . . .	109
<b>C</b>	<b>Results obtained by analyzing the hyperspectral EL image</b> . . . . .	<b>112</b>

## CONTENTS

---

C.1	Preprocessing . . . . .	112
C.2	Pipeline 1: PCA . . . . .	112
C.3	Pipeline 2: PCA + MCR . . . . .	115
C.3.1	Using 2 MCR components (without regularization) . . . . .	115
C.3.2	Using 3 MCR components (without regularization) . . . . .	116
C.3.3	Using 4 MCR components (without regularization) . . . . .	117
<b>D</b>	<b>XRF supplementaries</b>	<b>118</b>
<b>E</b>	<b>Supplementary material to Section 5.2</b>	<b>120</b>
<b>F</b>	<b>Supplementary material to Section 5.7</b>	<b>121</b>
<b>G</b>	<b>Supplementary material to Section 5.8</b>	<b>123</b>
G.1	Motivating regularized regression in MCR . . . . .	123
G.2	"Lack of fit" as a function of regularization strength . . . . .	124
G.3	Pure spectra and concentrations for the selected MCR models with three and four components . . . . .	125
G.4	Score images for MCR model with five components and $\alpha = 0.15$	127
G.5	Four-component MCR models with different regularization strengths . . . . .	128
G.6	Decompositions and reconstructions of mean PL spectra for re- gions A, B, C, and D . . . . .	128

**List of Figures**

2.1	Direct vs. indirect bandgap semiconductors . . . . .	6
2.2	Direct (band-to-band) recombination . . . . .	7
2.3	Shockley-Read-Hall recombination . . . . .	8
2.4	Auger recombination . . . . .	9
2.5	Cubic perovskite crystal structure . . . . .	10
2.6	Methylammonium lead iodide's (MAPbI <sub>3</sub> ) electronic dispersion diagram . . . . .	12
2.7	Light-induced phase segregation in a mixed-halide perovskite solar cell . . . . .	13
2.8	Planar regular (n-i-p) perovskite solar cell . . . . .	15
2.9	Energy band diagram of an example perovskite solar cell . . . . .	16
2.10	Hyperspectral data cube (hypercube) . . . . .	19
2.11	Image acquisition using a line scan/push-broom camera . . . . .	20
2.12	Inner workings of a line scan/push-broom camera . . . . .	21
3.1	Hyperspectral photoluminescence imaging setup . . . . .	30
4.1	Flowchart representation of the data analysis pipelines. . . . .	40
4.2	Method for initializing MCR . . . . .	42
5.1	Bandgap as a function of the halide stoichiometry. . . . .	46
5.2	Evolution of PL line spectrum . . . . .	47
5.3	Total PL intensity map of sample 1 . . . . .	49
5.4	Mean PL spectrum from the hyperspectral image of sample 1 . . . . .	50
5.5	Wavelength band corresponding to the PL maximum of the sample 1 image . . . . .	51
5.6	Comparing PL intensities from samples 1 and 3. . . . .	52
5.7	Elemental maps obtained by performing XRF imaging of sample 1. . . . .	54
5.8	Counts in the elemental maps from Figure 5.7 as a function of position along a line. . . . .	55
5.9	Effect of the background noise correction scheme. . . . .	57
5.10	Normalization constants (integrated PL intensities) for sample 1 . . . . .	58
5.11	Mean PL spectrum in the energy domain (sample 1) . . . . .	59
5.12	Scree plot for PCA performed on the non-normalized sample 1 image . . . . .	60
5.13	PCA scores and loadings (sample 1, without normalization) . . . . .	61
5.14	Interpretation of PC1 (without normalization) . . . . .	62
5.15	Scree plot for PCA performed on the normalized sample 1 image . . . . .	64
5.16	PCA scores and loadings (sample 1, with normalization) . . . . .	66
5.17	Regions with high or low PC2 scores (without normalization) . . . . .	68

LIST OF FIGURES

---

5.18	Pair-wise comparisons of the mean PL spectra of the selected regions from Figure 5.17. . . . .	69
5.19	"Lack of fit" ( <i>LoF</i> ) as a function of the number of MCR components (without regularization) . . . . .	72
5.20	Pure spectra of non-regularized MCR models . . . . .	73
5.21	"Lack of fit" ( <i>LoF</i> ) as a function of the number of MCR components for different values of the regularization parameter $\alpha$ . . . . .	75
5.22	Pure spectra of MCR models with three, four, and five components, both with and without regularization . . . . .	78
5.23	Pixel-wise "lack of fit" ( <i>PLoF</i> ) maps and histograms for the non-regularized three-component model and the regularized four-component model. . . . .	80
5.24	Comparing the decompositions and reconstructions of region C's mean PL spectra obtained using the MCR models with three and four components. . . . .	81
5.25	Thresholded pixel-wise "lack of fit" ( <i>PLoF</i> ) map (three MCR components) . . . . .	82
5.26	Comparison of how the models with three and four MCR components decompose and reconstruct selected pixel spectra from the circular region in Figure 5.25. . . . .	83
5.27	Pure spectra for the MCR model with four components and $\alpha = 0.15$ . . . . .	85
C.1	Normalization constants (EL image) . . . . .	112
C.2	Scree plot (EL image) . . . . .	112
C.3	PCA scores and loadings (EL image) . . . . .	113
C.4	PCA scatter plots (EL image) . . . . .	114
C.5	MCR with two components (EL image) . . . . .	115
C.6	MCR with three components (EL image) . . . . .	116
C.7	MCR with four components (EL image) . . . . .	117
D.1	XRF settings. . . . .	118
D.2	XRF spectrum for sample 1. . . . .	119
D.3	XRF concentrations for sample 1. . . . .	119
E.1	Spectrally accumulated time series image of sample 2. . . . .	120
F.1	Selected pixel spectra from region D in Figure 5.17. . . . .	121
G.1	Selected score images and the pure components of a non-regularized MCR model with six components . . . . .	123
G.2	"Lack of fit" ( <i>LoF</i> ) as a function of the regularization parameter $\alpha$ for MCR models with different numbers of components. . . . .	124



## LIST OF FIGURES

---

G.3 MCR model with three components and $\alpha = 0$ (PL image, sample 1) . . . . .	125
G.4 MCR model with four components and $\alpha = 0.15$ (PL image, sample 1) . . . . .	126
G.5 Score images for the MCR model with five components and $\alpha = 0.15$ . . . . .	127
G.6 Pure spectra of MCR models with four components and different regularization strengths. . . . .	128
G.7 Decompositions and reconstructions of the mean PL spectra for regions A, B, C, and D. . . . .	129

**List of Tables**

1	Description of PSC samples . . . . .	32
2	Overview of HSPL images and their respective settings . . . . .	33
3	Locations of the pure spectra's PL intensity peaks and the corresponding bromide-to-iodide ratios $y$ (MCR model with four components) . . . . .	86
4	Locations of the pure spectra's PL intensity peaks and the corresponding bromide-to-iodide ratios $y$ (MCR model with three components) . . . . .	87

---

# 1 Introduction

In 2019, the European Commission presented its *European Green Deal* that aims to reduce greenhouse gas emissions by at least 55% by 2030, compared to the 1990 levels. This is one key milestone towards reaching net-zero emissions in Europe by 2050 [1]. Achieving these targets will require a large-scale transition from fossil fuels to renewable energy resources [2]. One technology that is set to play an important role in this transition is solar photovoltaic (PV) systems. According to the International Energy Agency (IEA), the electricity generation from PV systems will surpass the generation from wind power in 2028 and become the largest contributor to the global electricity generation among the variable renewable technologies [3].

To reach higher power conversion efficiencies from PV systems, vast amounts of research need to be put into exploring novel PV materials and designs. High power conversion efficiencies are crucial for minimizing land use and cost per unit of energy produced by PV systems [4]. Currently, crystalline silicon-based PV modules, particularly those made from monocrystalline silicon, dominate the PV market [5]. Zanatta et al. [6] state that an updated version of the Shockley-Queisser limit gives single-junction crystalline silicon solar cells a maximum efficiency of 33%. However, other researchers claim that Auger recombination limits their efficiencies to below 29.4% [7–9]. State-of-the-art silicon-based solar cells are already relatively close to this number. According to NREL’s *Best Research-Cell Efficiencies* chart [10], a *heterojunction with intrinsic layer* (HIT) cell holds the record efficiency among the non-concentrating silicon cells. The cell, which was manufactured by LONGi, had an efficiency of 27.1%. Furthermore, the efficiency of non-concentrating silicon solar cells has only increased by 0.5 percentage points since 2016, and it seems to have plateaued [10]. Thus, the efficiency growth for commercialized silicon solar cells is also expected to slow down [4]. These factors highlight the need for novel solar cell designs and materials to enable power conversion efficiencies to increase by more than a few percentage points.

Hybrid organic-inorganic lead halide perovskites have gained interest in the scientific community in recent years, due to material properties that make them suitable for use in solar cells [11]. They have high absorption coefficients, tunable bandgaps, long carrier lifetimes, and long carrier diffusion lengths [11, 12]. Moreover, perovskite absorber layers can be deposited on substrates by using relatively simple, inexpensive fabrication methods [4]. This, in addition to the aforementioned optoelectronic properties, makes them, for instance,

---

suitable for use as top cells in perovskite-silicon tandem cells [4]. Tandem configurations can achieve even higher efficiencies than single-junction devices, and power conversion efficiencies exceeding 30% are demonstrated for small-area versions of these perovskite-silicon tandem devices [4].

Materials having a perovskite structure have been known for a long time, but it was not until recently they were introduced to photovoltaic applications [11]. In 1958, Christian Møller published his work on the crystal structure of cesium lead halides [13]. In the work leading up to his article, Møller discovered that these materials had a perovskite structure. Moreover, because of their colors, he stated that “one might guess that the former crystals also have special electrical properties” [13]. Another important step towards perovskite solar cells was taken 20 years later when Dieter Weber introduced organic-inorganic hybrid perovskites [11]. He synthesized the first methylammonium lead halide perovskites, in which the methylammonium ion ( $\text{CH}_3\text{NH}_3^+$ ) replaced the cesium ion ( $\text{Cs}^+$ ) [14]. However, it would take another 30 years before perovskites were introduced to solar cell applications. Perovskites were first introduced to photovoltaic applications by researchers working on dye-sensitized solar cells approximately 15 years ago [15]. In the following years, various perovskite solar cell architectures were investigated, before Liu et al. [16] demonstrated that it was possible to obtain high power conversion efficiencies using simple planar structures [11].

Nevertheless, perovskite solar cells (PSCs), both single-junction and tandem devices, have some obstacles to overcome before large-scale commercialization can become a reality. Their long-term stability is the main concern [4]. Hidalgo et al. [12] point out that outdoor application and commercialization depend on a greater understanding of the degradation mechanisms. Perovskites are, for instance, sensitive to temperature, oxygen, and UV light [12].

One important instability mechanism in perovskite solar cells is light-induced phase segregation, which means that domains of higher and lower concentrations of the perovskite’s halide species are formed upon light exposure [4]. Phase segregation represents a loss mechanism in the solar cell because it induces lower bandgap domains that act as charge carrier traps [17]. These trap states have negative effects on charge transport in the cell [4]. Moreover, local bandgap changes caused by phase segregation reduce the stability of tandem devices by causing a current mismatch between the two subcells [4].

Imaging characterization tools are vital for developing stable and highly efficient perovskite solar cells, since they allow researchers to gain a wider understanding of the optical, electronic, chemical, and physical properties of

perovskites [12, 18]. Techniques such as scanning electron microscopy (SEM) and transmission electron microscopy (TEM) can be used to study the morphology of samples, whereas optoelectronic properties can be examined via photoluminescence (PL) mapping [12]. These imaging techniques have facilitated important advances in perovskite solar cell technology. SEM images have, for instance, been used to improve the perovskite coverage after spin-coating deposition [19]. SEM images have also been used to demonstrate that the morphology of perovskite layers can be improved by using a flow of gas during the spin-coating deposition [20].

More recently, hyperspectral photoluminescence (HSPL) imaging has been utilized to study perovskite solar cells. Nguyen et al. [18] used HSPL imaging to map various optoelectronic properties of single-junction perovskite solar cells, such as the optical bandgap, absorptivity and implied open-circuit voltage. This technique allowed them to study the correlations among these parameters, in addition to examining how their spatial variation evolved after degradation tests. By examining the bandgap changes after degradation tests, they were able to confirm phase segregation in their samples [18].

### 1.1 Objective and Project Description

The main goal of this master's thesis is to use HSPL imaging, which is a contactless and non-destructive imaging technique [18], to map light-induced phase segregation in the perovskite layers of mixed-halide hybrid perovskite solar cells. The present study focuses on testing and developing data analysis techniques that can be used to extract relevant information about phase segregation from the hyperspectral images. These data analysis techniques include principal component analysis (PCA) and multivariate curve resolution (MCR), in addition to necessary preprocessing steps. MCR has previously been used for extracting defect-related PL signals from hyperspectral images of silicon solar cells [21, 22]. Furthermore, HSPL images have, as previously mentioned, been used to detect phase segregation in perovskite solar cells. However, no known studies have focused on the spatial and spectral information that can be extracted by using the aforementioned data analysis techniques or how they can be used for mapping phase segregation in mixed-halide perovskites.

To investigate these questions, planar n-i-p architecture perovskite solar cells are fabricated in REALTEK's laboratories by following the procedures described by Saliba et al. [23] and Tay et al. [24]. HSPL images of these cells are then acquired and analyzed using PCA and MCR. An integral part of the

work revolves around building an MCR model that is capable of extracting relevant signals from the HSPL images. These signals will be compared with results from other studies to examine whether they originate from phase segregation. X-ray Fluorescence (XRF) images of the samples are also acquired in an attempt to validate the results obtained by analyzing the HSPL images. Moreover, in addition to capturing HSPL images of the whole sample area, a time-series image of a single line on the sample is recorded to allow investigations of how light soaking affects the evolution of PL signals from the perovskite.

## 1.2 Research questions

The main research questions investigated in this work are:

- Does light-induced phase segregation occur in the perovskite samples studied in this work? Which indications do the HSPL images give?
- How can HSPL imaging and accompanying data processing steps be used to map light-induced phase segregation in mixed-halide hybrid perovskite solar cells?
- Can the pixels' PL spectra be utilized to predict regions of higher and lower concentrations of the halides in the perovskite absorbers?

Moreover, some additional questions were important for guiding the work on the data processing techniques:

- Which preprocessing steps are necessary?
- What information relevant for phase segregation can be extracted from the hyperspectral images using PCA and MCR?
- What are the optimal parameter settings regarding the aforementioned purposes?
- How can PCA be used to initialize the MCR algorithm?
- How may an MCR model that can decompose the pixel spectra into their constituents be built? How many components should be used?
- Which signal components, if any, originate from phase segregation?

---

## 2 Theory

The opening of this chapter, Section 2.1, introduces fundamental concepts related to solar cell and semiconductor physics such as energy bands, the distinction between direct and indirect bandgaps, and recombination mechanisms. Next, Section 2.2 introduces perovskites. The structure and optoelectronic properties of this family of materials are presented before the concept of phase segregation is discussed. Thereafter, in Section 2.3, the perovskite solar cell architecture and working principles are explained. The subsequent sections focus on the imaging and characterization techniques – hyperspectral photoluminescence (HSPL) imaging in Section 2.4, and X-ray fluorescence (XRF) imaging in Section 2.5. A discussion of the theoretical foundation and working principles behind the data analysis techniques being applied to the hyperspectral images concludes the chapter in Section 2.6.

### 2.1 Solar Cell and Semiconductor Fundamentals

All the fundamental solar cell and semiconductor physics principles presented in Section 2.1 are based on chapters 1, 3, 6, 7, 10, and 13 in *Solar energy: The physics and engineering of photovoltaic conversion, technologies and systems* by Smets et al. [8], unless indicated otherwise by citations to other sources.

A *photovoltaic (PV) cell*, or simply a *solar cell*, is a semiconductor-based device capable of converting incoming sunlight into electrical energy. In the semiconductor material, the allowed energy levels that electrons can occupy form so-called *energy bands*. These bands are separated by energy ranges without allowed energy states, called *bandgaps*. The bands determining the flow of electrical current in a semiconductor are those related to the valence electrons, namely the *valence band* and the *conduction band*. The energy difference between the maximum energy  $E_V$  of the valence band and the minimum energy  $E_C$  of the conduction band is referred to as the *bandgap energy*  $E_G$ . A photon having an energy greater than the bandgap energy of the semiconductor can be absorbed by the material. The photon then transfers its energy to an electron in the valence band and thus excites it to the conduction band. This leaves behind a *hole* in the valence band. The solar cell can deliver electrical energy to an external circuit if electrons and holes are separated within the device and extracted at its terminals.

Solar cells can either be single- or multi-junction devices. Single-junction devices are the simplest solar cells, having one photovoltaic absorber and, hence,

one bandgap. Multi-junction or tandem devices are essentially stacks consisting of several solar cells with different bandgaps [25]. These devices are designed for better utilization of the solar spectrum.

### 2.1.1 Direct and indirect bandgap semiconductors

If the maximum energy of the valence band and the minimum energy of the conduction band are located at the same crystal momentum value in the energy-momentum space of the electrons, the material is said to be a *direct bandgap semiconductor*. However, when the maximum of the valence band and the minimum of the conduction band correspond to different momentum values, the material is an *indirect bandgap semiconductor*. To excite an electron in a direct bandgap semiconductor, it is sufficient for the material to absorb a photon with an energy greater than the bandgap energy. On the other hand, an indirect bandgap also requires the electron to change its momentum, which can be fulfilled through an interaction with a phonon. A phonon is a quasiparticle describing the quantized vibrational modes in a lattice. Figure 2.1 illustrates the excitation process in direct bandgap and indirect bandgap semiconductors. Since direct bandgap semiconductors only require interactions with photons, their absorption coefficients are larger than those of indirect bandgap semiconductors. Thin-film solar cell materials like gallium arsenide (GaAs) and cadmium telluride (CdTe) are examples of direct bandgap semiconductors. Crystalline silicon (c-Si) is an example of an indirect bandgap semiconductor.

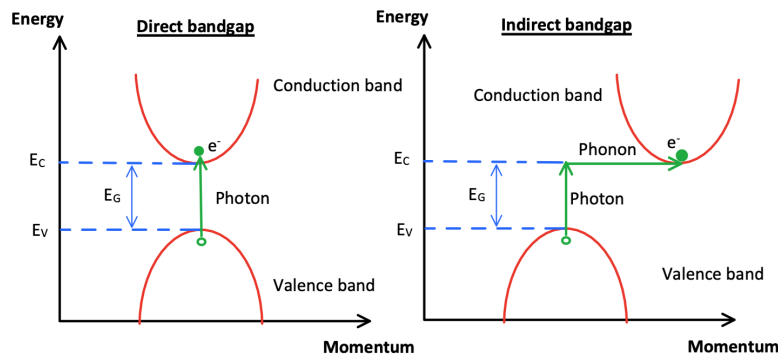


Figure 2.1: Schematic illustration of the excitation of an electron ( $e^-$ ) from the valence band to the conduction band in a direct bandgap and an indirect bandgap semiconductor. In both cases, a photon with an energy larger than the bandgap energy  $E_G$  is absorbed to provide the electron with energy. In an indirect semiconductor, the electron must also change its momentum by interacting with a phonon. The figure is created based on information found in [8, p. 66].



### 2.1.2 Recombination mechanisms

Recombination in a semiconductor means that an electron in the conduction band is deexcited to a lower energy state by filling a hole in the valence band. How the energy difference between the electron's initial and final state is released distinguishes different recombination mechanisms. *Radiative recombination* means that the energy is emitted as a photon, while *non-radiative recombination* means that other electrons, holes, or phonons receive this energy. The following text will briefly describe the three fundamental recombination mechanisms in solar cells.

First, *direct recombination*, or *band-to-band (BB) recombination*, refers to the process where an electron in the conduction band falls directly back to the valence band and recombines with a hole. The electron thus makes a direct transition between the bands, and the bandgap energy is usually emitted as a photon. Since no change in momentum is needed for a transition between the bands in a direct bandgap semiconductor, direct recombination is more efficient in these devices than in indirect bandgap semiconductors. Figure 2.2 illustrates direct recombination in a direct bandgap semiconductor.

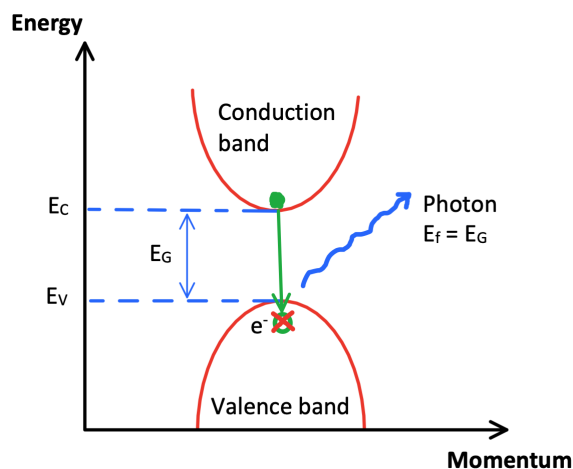


Figure 2.2: *Direct (band-to-band) recombination in a direct bandgap semiconductor. An electron from the conduction band recombines with a hole in the valence band, and a photon with energy  $E_f$  equal to the bandgap energy  $E_G$  is emitted. The figure is created based on information found in [8, p. 67].*

Second, *Shockley-Read-Hall (SRH) recombination* is recombination facilitated by impurities or lattice defects. Vacancies or foreign atoms in the semiconductor lattice are examples of such material imperfections that may promote recombination by introducing trap states within the forbidden bandgap [21].

The trap states may capture electrons from the conduction band and consequently facilitate recombination with holes from the valence band. SRH recombination can either be radiative by emitting photons or non-radiative by converting the excess energy to phonons so that it is distributed as heat [26]. SRH recombination through one or several trap states is illustrated in Figure 2.3.

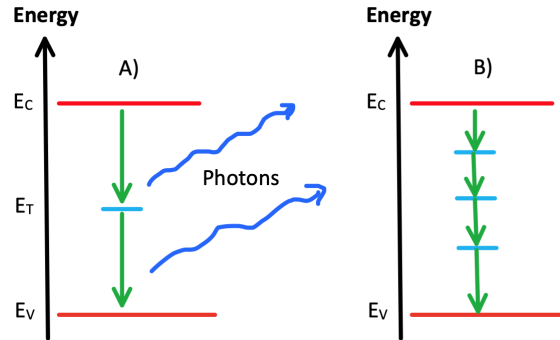


Figure 2.3: *Schematic illustration of Shockley-Read-Hall (SRH) recombination.  $E_T$  indicates the energy level of a trap state. The process can either be radiative, as indicated in A), or non-radiative as in B). In A), photons with energies lower than the bandgap energy are emitted. The recombination process may occur through one or more defect levels, as shown in A) and B), respectively. The illustration is based on a figure from [22].*

Third, *Auger recombination* is a non-radiative recombination mechanism in which a third charge carrier, an electron or a hole, receives energy from the recombining electron-hole pair. Subsequently, the excited charge carrier transfers its energy to vibrations in the semiconductor lattice. Auger recombination becomes more important as doping levels increase and in situations with large concentrations of free charge carriers. A simple illustration of Auger recombination is given in Figure 2.4.

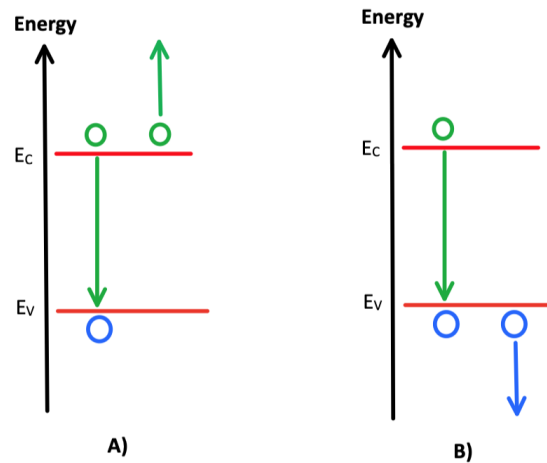


Figure 2.4: *Schematic illustration of Auger recombination. Electrons are represented by green circles, and holes are represented by blue circles. The excess energy from the annihilation of the electron-hole pair can either be transferred to an electron as in A), or to a hole as in B). The illustration is based on a figure from [8, p. 76], but with slight modifications for clarity.*

Recombination is important for solar cell performance, as well as for characterization techniques. Regarding solar cell performance, it is important to be able to extract charge carriers before they recombine. The reason is that charge carriers that recombine before they are collected at the terminals represent a loss in the solar cell, which leads to a decrease in its efficiency. Therefore, it is important to, for instance, reduce the density of trap states by having as few impurities and defects as possible. When it comes to characterization techniques, both BB and defect-related recombination (SRH) may produce luminescence signals that can be used to study semiconductors [21]. This will be discussed in greater detail in Section 2.4.1.

## 2.2 Perovskites

### 2.2.1 Perovskite structure

The name *perovskite* originates from a mineral with the same name, having the chemical formula  $\text{CaTiO}_3$ . It is named after the mineralogist Lev A. Perovski [8, p. 210]. General perovskites can be described by the chemical formula  $\text{ABX}_3$ . Here, A is an organic or inorganic cation, B is a metal cation, and X is a halide [11]. Ideal perovskite crystals have a cubic structure, as shown in Figure 2.5.

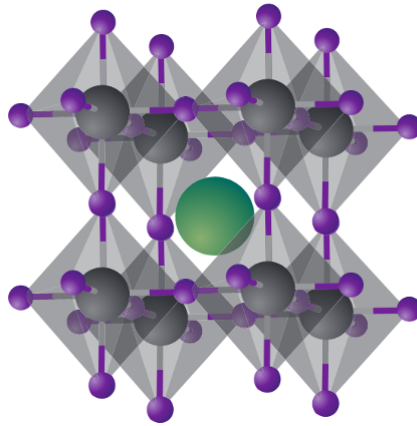


Figure 2.5: *Illustration of the general cubic perovskite structure. The green position is the organic or inorganic cation (A), the gray positions represent the metal cations (B), and the purple positions represent the halides (X). Reprinted with permission from [11].*

Perovskites for photovoltaic applications are often organic-inorganic compounds [8, p. 210], referred to as *hybrid organic-inorganic perovskites* [27]. A commonly used A-site cation is methylammonium (MA), with the chemical formula  $\text{CH}_3\text{NH}_3^+$  [8, p. 210]. Other options include formamidinium (FA), cesium (Cs), and rubidium (Rb) ions [11, 23]. The B-site cation is typically lead (Pb) [8, p. 210]. This may be problematic, because of toxicity [11]. An alternative to lead is tin (Sn), but tin-based perovskites are highly sensitive to oxygen [11], and thus less stable [8, p. 211]. The halides in position X are typically chloride (Cl), bromide (Br), or iodide (I) [11]. Perovskites containing more than one X-site halide are generally referred to as *mixed-halide perovskites* in the literature.

### 2.2.2 Bandgap and optoelectronic properties

Although the progress in perovskite solar cell efficiencies has been astonishing, it has remained unclear whether a direct or an indirect bandgap model should be used to explain the excellent optoelectronic properties of hybrid organic-inorganic perovskites. Traditionally, perovskites have been classified as direct bandgap semiconductors [28]. Hybrid perovskites combine large absorption coefficients and long charge carrier lifetimes [29], which makes them ideal for solar cell applications. However, in conventional semiconductors, there is usually a trade-off between strong absorption and long carrier lifetimes, as a consequence of the material's bandgap. Direct bandgap semiconductors have large absorption coefficients and short charge carrier lifetimes, whereas indirect bandgap semiconductors show weak absorption and long lifetimes [29].

That hybrid perovskites have an indirect-direct bandgap, with an indirect transition slightly below the direct transition, is one possible explanation of the aforementioned optoelectronic properties. For example, Wang et al. [29] state that methylammonium lead iodide ( $\text{MAPbI}_3$ ) perovskite has an indirect bandgap transition 60 meV below the direct bandgap transition. The band diagram of this perovskite is shown in Figure 2.6. The direct transition gives efficient absorption of incoming photons. However, since the excited charge carriers thermalize to regions in the conduction band at a different crystal momentum, an indirect transition is needed for the electrons to recombine with the holes [29]. Indirect transitions occur less frequently since they require a change in crystal momentum [27], and the charge carrier lifetimes will hence be prolonged by this indirect bandgap transition. The indirect transition is caused by *spin-orbit coupling*, resulting in so-called *Rashba-splitting* of the conduction band [29].

An important feature of lead halide perovskite bandgaps is that they can be tuned by altering the halide stoichiometry. For instance, Kulkarni et al. [31] were able to tune the bandgap of methylammonium-based lead halide perovskites from approximately 1.56 eV to 2.23 eV by increasing the bromide-to-iodide ratio. The cation at the A-site in Figure 2.5 will also influence the bandgap, and Li et al. [32] have developed machine-learning models for predicting the bandgaps of lead halide perovskites based on their chemical compositions. Their linear regression model predicts, in agreement with Kulkarni et al. [31], that the bandgap increases as the bromide fraction increases. The fact that metal halide perovskite bandgaps can be tuned within a quite broad range is one of the factors that makes them suitable for tandem applications [33].

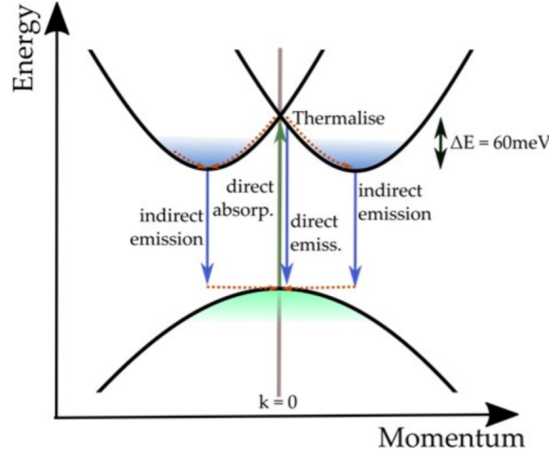


Figure 2.6: *Electronic dispersion diagram showing the valence band and conduction band in methylammonium lead iodide (MAPbI<sub>3</sub>). The symbol  $k$  represents the wave vector, a physical quantity related to momentum [8]. At  $k = 0$ , the material has a direct bandgap transition. However, an indirect bandgap transition lies 60 meV below the direct bandgap transition. Thermalization down the conduction band valleys limits recombination, as a change in momentum is required before recombination. Figure reprinted with permission from *Advances in Engineering* [30].*

The linear regression model from Li et al. [32] predicts the bandgaps of perovskites with the chemical formula  $\text{Cs}_a\text{FA}_b\text{MA}_{(1-a-b)}\text{Pb}(\text{Cl}_x\text{Br}_y\text{I}_{(1-x-y)})_3$ . The bandgap  $E_G$  (in eV) is calculated as

$$E_G = -4.960 + 2.214a - 0.315b + 0.814x + 0.436y + 4.913R. \quad (2.1)$$

Here,  $a$ ,  $b$ ,  $x$ , and  $y$  are the stoichiometric coefficients defined in the chemical formula above.  $R$  is a factor that takes the radii of the ions into account. The factor is defined as

$$R = \frac{ar_{\text{Cs}} + br_{\text{FA}} + (1 - a - b)r_{\text{MA}}}{xr_{\text{Cl}} + yr_{\text{Br}} + (1 - x - y)r_{\text{I}}}, \quad (2.2)$$

where  $r$  represents the Shannon radii of the ions. They have the values  $r_{\text{Cs}} = 1.81 \text{ \AA}$ ,  $r_{\text{FA}} = 2.79 \text{ \AA}$ ,  $r_{\text{MA}} = 2.70 \text{ \AA}$ ,  $r_{\text{Cl}} = 1.81 \text{ \AA}$ ,  $r_{\text{Br}} = 1.96 \text{ \AA}$ , and  $r_{\text{I}} = 2.03 \text{ \AA}$  [32].

### 2.2.3 Light-induced phase segregation

Light-induced phase segregation in a perovskite material (also referred to as the *Hoke effect*) means that domains of higher and lower concentrations of the perovskite’s halide species are formed upon light exposure [34]. However, when the perovskite material is allowed to relax in the dark, it returns to a homogeneous, unsegregated state [35]. Figure 2.7 shows an illustration of phase segregation in a mixed-halide perovskite containing iodide and bromide.

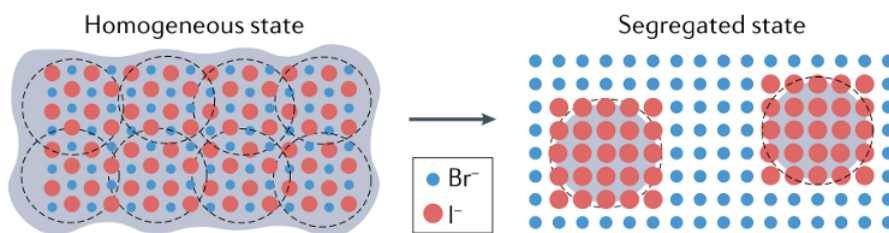


Figure 2.7: *Illustration of phase segregation in a mixed-halide perovskite. Left: a homogeneous state where bromide and iodide ions are uniformly distributed throughout the material. Right: a segregated state. The perovskite now contains separate iodide- and bromide-rich domains. Reprinted with permission from [4].*

As was mentioned in Section 2.2.2, increasing the bromide-to-iodide ratio in a mixed-halide perovskite leads to an increase in the perovskite’s bandgap. However, by studying  $\text{MAPb}(\text{Br}_x\text{I}_{1-x})_3$  perovskites with different values of  $x$ , Noh et al. [36] demonstrated that the increase in the bandgap caused by increasing  $x$  was not necessarily accompanied by an increase in the open-circuit voltage [35]. Interestingly, the open-circuit voltage decreased substantially when  $x$  increased from 0.2 to 0.58 [36]. Hoke et al. [37] linked this loss in performance to light-induced phase segregation and hypothesized that it was a consequence of charge carriers being trapped in lower-bandgap, iodide-rich domains.

Chen et al. [17] proposed a unified theory for light-induced phase segregation, based on the Helmholtz free energy. They explained that since the halides are quite mobile, there will be fluctuations in the local halide stoichiometry in the perovskite film. Consequently, lower-bandgap iodide-rich domains may spontaneously form. Photoexcited charge carriers will accumulate in these regions, as this lowers their free energies. The free energy is reduced even more if these iodide-rich domains grow. The result is an inward diffusion of iodide and an outward diffusion of bromide that allow this growth [17]. Bischak

et al. [38] estimated that the iodide-rich domains formed in  $\text{MAPb}(\text{Br}_{0.9}\text{I}_{0.1})_3$  were approximately 8-10 nm in diameter.

In addition to decreasing the open-circuit voltage, phase segregation also has detrimental effects on perovskite solar cell performance through other mechanisms. First, trap states introduced by phase segregation both promote non-radiative recombination and inhibit charge transport in the cell [4]. Moreover, since recombination reduces the number of charge carriers that can be collected at the electrodes, light-induced phase segregation may also reduce the cell's short-circuit current [8, p. 135] [34]. Second, phase segregation leads to local bandgap variations, which can cause current mismatch in tandem cells [4].

As Duan et al. [4] state, researchers debate whether phase segregation influences long-term stability or not. On the one hand, it has been shown that phase segregation is reversible since storing a perovskite in the dark helps it return to the unsegregated state and recover its initial performance [34]. Moreover, it has been shown that phase segregation can be reversed through illumination at sufficiently high intensities [39]. Experiments have also been conducted by Hoke et al. [37] in which perovskites were cycled between the segregated and unsegregated states. The perovskites did not show any noticeable signs of degradation after this cycling [35]. Therefore, some researchers claim that phase segregation is more of a short-term than a long-term instability issue [4]. On the other hand, other researchers claim that phase segregation changes the perovskite structure and composition, and thus should be treated as a long-term stability concern [4].



## 2.3 Perovskite Solar Cells

### 2.3.1 Cell architecture

Single-junction perovskite solar cells (PSCs) can be fabricated in various architectures. Some typical cell architectures are the planar regular (n-i-p) and planar inverted (p-i-n) devices, in addition to mesoporous (n-i-p) structures [23]. The perovskite solar cells manufactured and studied in this work are planar regular (n-i-p) devices. The fabrication process will be explained in Section 3.1.

A cross-sectional illustration of a planar regular (n-i-p) perovskite solar cell is given in Figure 2.8. Starting from the bottom, the cell has a glass substrate with a transparent conductive oxide (TCO) on top of it. The TCO is the solar cell's front contact [8, p. 178]. An electron transporting layer (ETL) is the stack's next layer. On top of it, the perovskite absorber and a hole transporting layer (HTL) are deposited. The uppermost layer is a metal electrode that serves as the back contact [23]. In an inverted (p-i-n) structure, the ETL and HTL switch places [23]. The purpose of these layers is to separate electrons and holes so that they can be extracted at their respective terminals [8, p. 23].

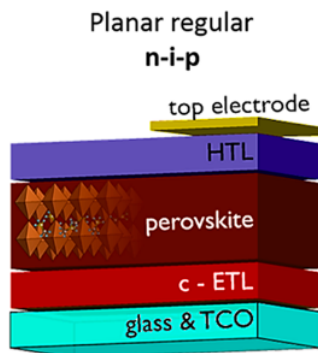


Figure 2.8: *Illustration showing how the different layers are stacked in a planar regular (n-i-p) perovskite solar cell. The bottom layer is a glass substrate with a transparent conductive oxide (TCO) on top of it. The TCO is the cell's front contact. The next layer is a compact electron transporting layer (c-ETL), in contrast to the mesoporous layer used in mesoporous architectures. The perovskite absorber layer and the hole transporting layer (HTL) are deposited on top of the ETL. The uppermost layer is a metal electrode (the back contact). Reprinted with permission from [23]. Copyright 2018 American Chemical Society.*

### 2.3.2 Working principles of the perovskite solar cells

An energy band diagram can be used to understand the basic principles of charge transport in a perovskite solar cell. Figure 2.9 shows (a) an example of a planar regular (n-i-p) perovskite solar cell, as well as (b) its energy band diagram. In this example, the back contact is made of gold (Au), and the TCO front contact is indium tin oxide (ITO) [40].

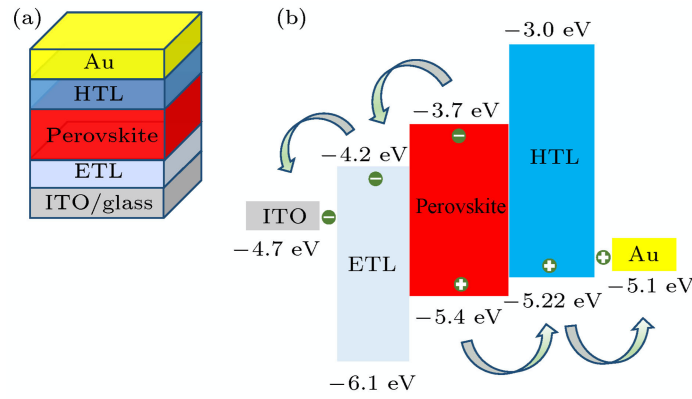


Figure 2.9: *Illustration of (a) the structure and (b) the energy band diagram of a planar regular (n-i-p) perovskite solar cell. The cell consists of a gold (Au) back contact, a hole transporting layer (HTL), a perovskite absorber, an electron transporting layer (ETL), an indium tin oxide (ITO) front contact, and a glass substrate. The energy values are relative to the vacuum energy level, namely the energy level where electrons are not influenced by the material anymore [8, p. 85] [41]. The values above the boxes represent the conduction band minimums, and the values below them are valence band maximums. Therefore, the boxes themselves and their heights represent the bandgaps of the ETL, the perovskite, and the HTL. The energies associated with ITO (-4.7 eV) and gold (-5.1 eV) are the work functions of these electrodes [40]. Work functions are related to the energy required to remove electrons from a material [8, p. 100]. The figure is reprinted with permission from [40].*

The arrows in Figure 2.9 (b) indicate in which direction electrons and holes move in the solar cell. The motion of these particles can be understood by examining the energy levels in the band diagram. First, free electrons and holes are generated when light is absorbed in the perovskite layer. Diffusion is then responsible for bringing the charge carriers to the interfaces between the perovskite absorber and the charge-transporting layers [42]. An electron in the perovskite's conduction band has an energy of around -3.7 eV relative to the vacuum level and, thus, enough energy to move into the conduction band of the ETL. However, since -3.7 eV lies in the middle of the HTL's bandgap, the

electron does not have enough energy to move into the conduction band of the HTL. This way, electrons will move down the conduction band slope to the ITO contact. Holes, on the other hand, will move up the valence band slope to the gold contact [8, p. 192]. Electrons and holes can then be extracted at their respective electrical contacts so that electrical energy is delivered to an external load [8, p. 23].

## 2.4 Hyperspectral Photoluminescence (HSPL) Imaging

Since absorption of incoming light and the subsequent excitation of electron-hole pairs is at the core of solar cell operation, different spectroscopy techniques are suitable for understanding the optical and electrical properties of solar cell materials [12]. One such technique is *photoluminescence* (PL) spectroscopy. In this work, PL spectroscopy is combined with hyperspectral imaging in a technique referred to as *hyperspectral photoluminescence* (HSPL) imaging [21, 43]. HSPL imaging allows studying the electrical and optical properties of perovskite solar cells by providing spatially and spectrally resolved PL images of the samples [18]. In the following subsections, the basic principles of photoluminescence and hyperspectral imaging will be presented.

### 2.4.1 Photoluminescence (PL) and electroluminescence (EL)

*Photoluminescence* (PL) refers to the emission of light from a material after an external illumination source has been used to excite it [44, p. 13]. First, incoming photons with energies larger than the material's bandgap excite electrons from the valence band to the conduction band [12]. Then, radiative recombination caused by either BB or SRH processes yields luminescence signals (emitted photons) that can be detected [12, 21]. To excite the electrons, one may use a laser having photon energies larger than the sample's bandgap energy [21].

*Electroluminescence* (EL) is another luminescence phenomenon used to study solar cell materials. EL is similar to PL, except for the fact that the excitation is achieved by connecting an electrical power source to the solar cell. This way, charge carriers that may recombine are injected into the material [44].

#### **Factors influencing EL/PL intensities: defects and film thickness**

Various factors influence the intensities of the emitted EL/PL signals. This section discusses two such factors: film thickness and material defects.

First, the thickness of the perovskite absorber layer will influence the intensities

of luminescence signals since it is related to the absorptivity of the absorber. The generalized Planck's emission law shows that the flux of luminescent photons is proportional to the absorptivity of the sample [45, 46]. This can be understood intuitively for PL signals by realizing that when more incident photons are absorbed, more electron-hole pairs are generated in the absorber layer. These excited charge carriers can recombine radiatively, yielding one luminescent photon per recombining electron-hole pair.

The absorptivity of the perovskite film depends on its thickness through the Lambert-Beer law. From this law, it follows that the light intensity  $I_{abs}(d)$  absorbed by an absorber layer of thickness  $d$  is given by the following equation:

$$I_{abs}(d) = I_0 [1 - e^{-\alpha d}]. \quad (2.3)$$

Here,  $\alpha$  is the material-specific absorption coefficient, and  $I_0$  is the incident light intensity [8, p. 140]. Therefore, spatial variations in luminescence intensities across the perovskite surface may result from variations in absorptivity caused by a non-uniform film thickness. A thicker absorber layer will absorb more light than a thinner layer and hence yield stronger luminescence intensities according to the previously mentioned generalized Planck's emission law.

Second, recall from the paragraph on SRH recombination in Section 2.1.2 that material defects introduce trap states within the band gap of the semiconductor. The defects influence luminescence intensities in two different ways, depending on whether the recombination through the trap states is radiative or not. First, if an electron-hole pair recombines non-radiatively, this happens at the expense of radiative recombination. Therefore, the intensity of the EL/PL signal is decreased. Second, if the electron-hole pair recombines radiatively, the emitted luminescent photon will have an energy lower than the bandgap energy, as can be seen in Figure 2.3. This weakens the EL/PL signal around the bandgap energy and produces an EL/PL signal at a lower energy level. As was mentioned in Section 2.2.3, phase segregation introduces trap states, and is hence one of the factors influencing the intensities of the PL-signals emitted from a perovskite solar cell.

### 2.4.2 Hyperspectral Imaging

The term *spectral imaging* refers to imaging using cameras capable of capturing light in several wavelength bands of the electromagnetic spectrum. A distinction is often made between *multispectral* imaging, in which a few, selected

wavelength bands are captured, and *hyperspectral* imaging, in which usually hundreds of contiguous wavelength bands are captured [47].

The acquired hyperspectral image is often presented as a hyperspectral data cube, or simply *hypercube*. This is a three-dimensional array representing the spatial dimensions  $X$  and  $Y$ , as well as the spectral dimension  $\lambda$ . Consequently, each pixel  $(X, Y)$  will have a spectrum of contiguous wavelengths [48]. These spectra contain valuable information about the physical properties of the sample being analyzed [18]. Figure 2.10 illustrates a hyperspectral data cube. For visualization purposes, the figure only includes six of the image's wavelength bands.

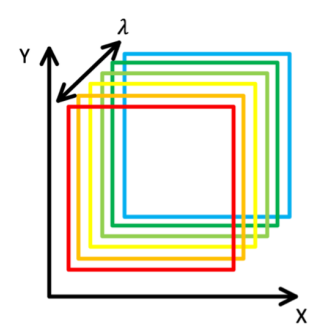


Figure 2.10: *Schematic illustration of a hyperspectral data cube (hypercube). The spatial dimensions are denoted by  $X$  and  $Y$ , and the spectral dimension is denoted by  $\lambda$ . Each colored rectangle represents a map of intensity values measured at different wavelengths. Each pixel  $(X, Y)$  will have a spectrum along the  $\lambda$ -dimension. Only six wavelengths are included in the illustration for visual clarity.*

### 2.4.3 Line scan/push-broom camera

Acquisition of the hyperspectral data cube can be achieved using different approaches. The camera used in this work is a *line scan* or *push-broom* camera. These cameras scan the sample line-by-line to obtain the three-dimensional hypercube. For each position of the camera, the spectra of a line of pixels are recorded [48]. In other words, one spatial dimension is acquired at a time, and the other spatial dimension is captured by moving the camera relative to the sample. This relative motion can be achieved by either moving the sample or by moving the camera [43]. The image acquisition using a line scan camera is illustrated in Figure 2.11.

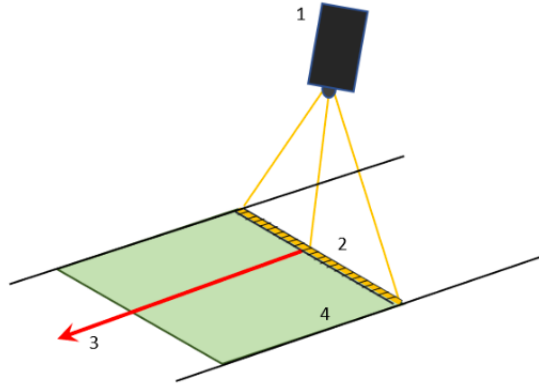


Figure 2.11: *Image acquisition using a line scan (push-broom) camera. The camera (1) obtains the hyperspectral data cube by scanning one line of pixels (2) at a time. It is possible to move either the camera (1) or the sample holder (4) to scan the sample in the direction indicated by the red arrow (3). The figure is reprinted from [43].*

The inner workings of a line scan camera are shown in Figure 2.12. First, light from a narrow line on the sample is allowed to enter the camera through the entrance slit [43]. After being collimated, the light is directed onto a dispersing element, namely a diffraction grating. Since the diffraction angle is wavelength-dependent, the different wavelengths can be separated [22]. The focusing optics focus the light onto a two-dimensional detector array, having a spatial dimension ( $X$ ) and a spectral dimension ( $\lambda$ ) [43]. The camera's detector array is usually a charge-coupled device (CCD) or a complementary metal-oxide semiconductor (CMOS) [49]. The second spatial dimension ( $Y$ ) is obtained by scanning new lines.

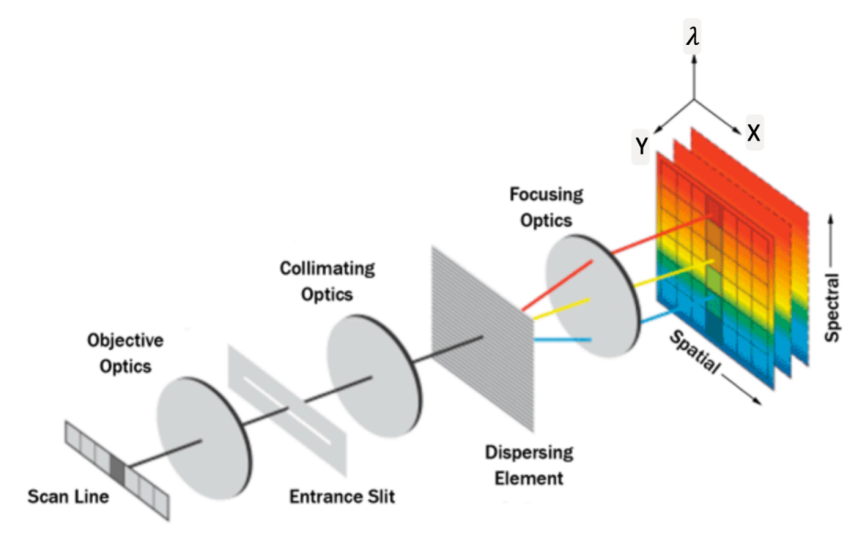


Figure 2.12: Working principles of a line scan (push-broom) camera. A narrow line of light is directed onto a dispersing element capable of separating the light into different wavelengths. Each scanned line gives a two-dimensional image, having one spatial dimension ( $X$ ) in addition to the spectral dimension ( $\lambda$ ). By scanning more lines, the second spatial dimension ( $Y$ ) is obtained. The figure is adapted with permission from [49].

## 2.5 X-ray Fluorescence (XRF) Imaging

X-ray fluorescence (XRF) imaging is a characterization technique that can, for instance, be used to acquire 2D maps showing the distribution of elements within a sample [50]. In XRF imaging, energetic X-rays are directed onto the sample. The working principle behind XRF is that inner-shell electrons are excited by these X-rays if the X-ray photons have energies larger than their binding energies [50]. The excitation creates inner-shell electron vacancies that can be filled by electrons from the outer shells [12, 51]. The energy difference between these two shells is released either via an Auger process (i.e. by ejecting an electron [50]) or by emitting a so-called fluorescent X-ray photon [52]. The emitted X-rays are specific to the element they originate from and can hence be used in XRF spectroscopy to identify different elements [51]. By examining the energies corresponding to peaks in the measured X-ray spectra, as well as the intensities of these peaks, the different elements and their respective concentrations can be identified and mapped [50].

## 2.6 Data Analysis

This section will present the theoretical foundations behind the techniques used to analyze the hyperspectral luminescence images, namely PCA and MCR. Before these techniques can be applied to the hyperspectral image, the hypercube has to be unfolded into a data matrix  $\mathbf{D}$  [48]. In this matrix, the rows represent the pixels and the columns their spectra. Using terminology from data science and machine learning, each pixel is therefore treated as an *observation* or *instance*, and each spectral band is a *feature* or *variable* [53]. After applying the data analysis algorithms, the results are refolded back into a three-dimensional data cube [48]. It is important to note that some data analysis packages, such as *Spectral Python*, automatically handle the unfolding and refolding processes. Other packages, for instance *pyMCR*, require the users to implement these steps themselves.

### 2.6.1 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a feature extraction technique that aims to extract the most relevant information from a dataset by constructing a new set of variables to describe it [53, p. 146]. These new variables, which are called *principal components* (PCs), are linear combinations of the original variables [54, p. 445]. The first principal component axis points in the direction of maximum variance in the dataset. The next principal components are constructed by finding directions of maximum variance under the constraint that the new principal component axes must be orthogonal to all previous principal component axes [53, p. 146]. This way, PCA allows the dataset to be described by a set of uncorrelated variables [53, p. 147].

PCA is performed by finding the eigenvalues and eigenvectors of the covariance matrix associated with the data matrix  $\mathbf{D}$  [53, p. 147]. The eigenvectors give the directions of the principal component axes and the eigenvalues quantify the variance explained by each principal component [53, p. 149] [54, p. 446]. The eigenpairs are sorted in decreasing order so that the eigenvector corresponding to the largest eigenvalue gives the first principal component (PC1), the eigenvector corresponding to the second largest eigenvalue gives the second principal component (PC2), and so on [53, p. 153].

Following the notation used by Amigo et al. [48], PCA can be expressed mathematically as



$$\mathbf{D} = \mathbf{T}\mathbf{P}^T + \mathbf{E}. \quad (2.4)$$

In this equation, the data matrix  $\mathbf{D}$  is decomposed into a score matrix  $\mathbf{T}$  and a loading matrix  $\mathbf{P}^T$ . The matrix  $\mathbf{E}$  represents the residuals [48].

The score matrix gives the coordinates of the observations (in this context pixels) relative to the principal component axes [55]. The loadings, which are given by the aforementioned eigenvectors, show how the original variables are weighted when forming the principal components [54, p. 445]. In the context of hyperspectral data, the loadings show how much each wavelength band contributes to the principal components. A high positive or negative value indicates that a wavelength band is important for a principal component [55].

One important application of PCA is dimensionality reduction, which means that the number of variables used to describe a dataset is reduced [53]. Regarding hyperspectral images, dimensionality reduction is achieved by using fewer principal components than the number of wavelength bands present in the original image. The rationale behind this is to use only the most relevant information, which is contained in the first principal components, for further analysis [53]. A *scree plot*, showing the variance explained by each component, can be used to determine how many principal components to use [54, p. 446] [56].

### 2.6.2 Multivariate Curve Resolution (MCR)

Multivariate Curve Resolution (MCR) is a technique used to decompose spectra into their constituents. The method assumes that each spectrum is a linear combination of a set of *pure spectra* [57]. Each of these pure spectra originates from the fundamental physical processes occurring in the system [22].

The decomposition is performed by solving the following equation:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E}. \quad (2.5)$$

Here,  $\mathbf{D}$  is the unfolded hypercube presented in Section 2.6, where the rows contain the individual pixel spectra.  $\mathbf{S}^T$  is the loading matrix, containing the pure spectra of the system [57]. The score matrix  $\mathbf{C}$  represents how each pure spectrum has to be weighted to reproduce the pixel spectra in the hyperspectral image [22]. These weights are normally referred to as *concentrations* [57]. The

reconstructions of the spectra in the data matrix  $\mathbf{D}$  are obtained by calculating the product  $\mathbf{C}\mathbf{S}^T$  between the concentrations and the pure spectra [58].

The matrix  $\mathbf{E}$  in Equation (2.5) is the residual matrix, representing factors not explained by the model, like noise and experimental errors [57, 59]. It represents the differences between the original spectra and their reconstructions, and it is calculated by subtracting  $\mathbf{C}\mathbf{S}^T$  from  $\mathbf{D}$ .

After Equation (2.5) is solved, the concentration matrix  $\mathbf{C}$  is reshaped back into a three-dimensional array to form 2D images showing the spatial variation in the concentrations of each pure component [57].

Equation (2.5) is solved iteratively by using an alternating regression (AR) scheme [59]. As input, the model needs initial guesses of either the system's concentrations or pure spectra [59]. With an initial estimate of either  $\mathbf{C}$  or  $\mathbf{S}^T$  in place, these matrices are updated iteratively by fixing one of them and performing regression on the other [59]. This way, the algorithm alternates between performing regression on  $\mathbf{C}$  and  $\mathbf{S}^T$ . One complete iteration includes updating both matrices. Therefore, an iteration is said to comprise two so-called *half-iterations*.

The regression aims to minimize the model's reconstruction error as measured by a selected *objective function* [59]. One such function is the *mean squared error* (*MSE*) between the original and reconstructed pixel spectra. The error  $MSE_{[k]}$  at half-iteration number  $k$  is calculated as follows:

$$MSE_{[k]} = \frac{\sum_{i,j} e_{[k],ij}^2}{M \cdot N}. \quad (2.6)$$

Here,  $e_{[k],ij}$  represents the element in the  $i$ th row and  $j$ th column of the residual matrix  $\mathbf{E}_{[k]}$  calculated at half-iteration number  $k$ .  $M$  and  $N$  are the number of rows and columns in  $\mathbf{E}_{[k]}$ , respectively [59]. The optimization process stops when a convergence criterion is met [58]. The convergence criteria are explained in greater detail in Section 4.3.2.

The regression can be customized in several ways by employing different objective functions, constraints on  $\mathbf{C}$  and  $\mathbf{S}^T$ , and regularization terms [59]. When working with hyperspectral images, it is common to use non-negative constraints on both matrices, since spectra and concentrations should have positive values from a physical point of view [57]. The simplest implementation of the AR algorithm uses ordinary least-squares (OLS) regression on  $\mathbf{C}$  and  $\mathbf{S}^T$ . It is referred to as *multivariate curve resolution-alternating least squares* (MCR-ALS) [59].

### Regularization and overfitting

Regularization is a useful technique for limiting the degree of *overfitting* in a regression model. Overfitting means that a model adapts too well to the given dataset, for instance by adapting to noise [53, p. 76]. A model suffering from overfitting is not capable of distinguishing the important information from the noise or randomness in the dataset. Overfitting can occur when a model has too many parameters (for instance too many MCR components), which results in a model that is too complex for the given dataset [53, p. 75].

Regularization means adding a penalization term to the regression model's objective function. One common form of regularized regression is called *Ridge Regression*. Ridge Regression includes an L2-term,

$$\alpha \cdot \sum_{j=1}^m w_j^2,$$

in the regression model's objective function [53, p. 338]. Here,  $w_j$  represents the regression model's weights. In MCR's AR scheme, these weights are the values in the  $\mathbf{C}$  and  $\mathbf{S}^T$  matrices. The parameter  $\alpha$  is referred to as the *regularization parameter*. The regularization strength can be varied by tuning this parameter [53, p. 338]. The effect of adding the L2-term above to the MCR model's objective function is that the model is penalized for having extreme weight values. This is a way to prevent it from becoming too complex [53, p. 337].

---

## 3 Experimental

This chapter describes the experimental work. First, Section 3.1 explains the fabrication of PSC samples. Next, the HSPL imaging setup is explained, and the image acquisition procedure is described in Section 3.2. Section 3.3 on the supplementary XRF imaging concludes the chapter.

### 3.1 Perovskite Solar Cell Fabrication

The PSCs examined in this study are planar regular (n-i-p) cells, prepared by following the procedures described by Saliba et al. [23] and Tay et al. [24]. The following section will provide an overview of the fabrication steps involved in the production of the cells. The fabrication steps described in Sections 3.1.3, 3.1.4, and 3.1.5 are performed inside a nitrogen-filled glovebox.

The samples studied in this work comprise all layers in Figure 2.8, except for the back contact (*top electrode*). Strictly speaking, these samples are not complete solar cells. However, they are referred to as *solar cells* throughout this work for simplicity.

#### 3.1.1 Substrate cleaning

Before other layers can be deposited, the substrates, which are 2.5 cm × 2.5 cm ITO-coated glass substrates, need to be cleaned. The cleaning procedure from Saliba et al. [23] is summarized below.

First, the substrates are brushed using a 2% Hellmanex solution. Then, they are submerged in the same Hellmanex solution in an ultrasonic bath. After 15 minutes in this bath, the samples are rinsed with deionized water. Subsequently, two 15-minute ultrasonic baths follow, first in isopropanol, and then in acetone. After these baths, the samples are rinsed with acetone and isopropanol, before a strong flow of nitrogen is applied to dry the isopropanol. The last cleaning step is UV-ozone cleaning for 15 minutes [23].

#### 3.1.2 Electron transport layer (ETL)

After cleaning the substrates, the next step is to deposit the SnO<sub>2</sub> (tin oxide) ETL, following the procedures described by Saliba et al. [23]. The SnO<sub>2</sub> layer is deposited using chemical bath deposition (CBD). The solution used for this purpose is prepared by first dissolving 0.5 g of urea in 400 mL of deionized

water. Thereafter, 10  $\mu\text{L}$  of mercaptoacetic acid and 0.5 mL of 37% HCl (hydrochloric acid) are added to the solution. The last step is to dissolve 0.1 g of  $\text{SnCl}_2 \times 2\text{H}_2\text{O}$  (tin(II) chloride dihydrate) in the solution. This solution is then stirred for at least two minutes before it is poured into a glass container together with the substrates.

The chemical bath deposition is carried out by heating the solution to 70 °C for three hours in a laboratory oven. Afterward, the substrates are placed in an ultrasonic bath with deionized water for two minutes. This is followed by an hour-long annealing process at 180 °C in a laboratory oven. The samples stay inside the oven overnight, which allows them to cool down slowly. Right before perovskite deposition, the  $\text{SnO}_2$  layer is cleaned using UV-ozone cleaning for 20 minutes [23].

#### 3.1.3 Perovskite precursor solution

The perovskite precursor solution is prepared following the procedures described by Tay et al. [24]. The solution is prepared by dissolving various salts in a 4:1 V/V dimethylformamide (DMF)/dimethyl sulfoxide (DMSO) solution. The following quantities of the salts are added to 1 mL of the solution:

- 18 mg CsI (cesium iodide)
- 25.4 mg MABr (methylammonium bromide)
- 84.7 mg  $\text{PbBr}_2$  (lead(II) bromide)
- 198.5 mg FAI (formamidinium iodide)
- 585.1 mg  $\text{PbI}_2$  (lead(II) iodide)

After dissolving the salts, the solution is heated to 50 °C for two hours on a hot plate [24].

By adding the aforementioned salts, the resulting perovskite layer will have  $\text{Cs}^+$ ,  $\text{MA}^+$ , and  $\text{FA}^+$  as A-site cations, and  $\text{Pb}^{2+}$  as the B-site metal cation. The halides at the X-site are  $\text{I}^-$  and  $\text{Br}^-$ . The chemical formula of the perovskite is calculated based on the amounts above, and it is approximately  $\text{Cs}_{0.048}\text{FA}_{0.80}\text{MA}_{0.15}\text{Pb}(\text{Br}_{0.15}\text{I}_{0.85})_3$ . This means that the coefficients in the general chemical formula  $\text{Cs}_a\text{FA}_b\text{MA}_{(1-a-b)}\text{Pb}(\text{Cl}_x\text{Br}_y\text{I}_{(1-x-y)})_3$  are  $a = 0.048$ ,  $b = 0.80$ ,  $x = 0$ , and  $y = 0.15$ .

### 3.1.4 Perovskite film deposition

The perovskite film is deposited on top of the ETL using a two-step spin-coating procedure described by Saliba et al. [23]. Before spinning, 120  $\mu\text{L}$  of the perovskite precursor solution is dropped onto the substrate. The first spin-coating step lasts for 10 s and involves spinning at 1000 rpm with 200 rpm/s acceleration. The second step lasts for 20 s and involves spinning at 6000 rpm with 2000 rpm/s acceleration. Moreover, 200  $\mu\text{L}$  of chlorobenzene (antisolvent) is dropped in the middle of the substrate 15 s into the second spin-coating step. Right after the spin-coating, an annealing step is conducted using a hot plate preheated to 100  $^{\circ}\text{C}$ . The substrates are moved from the spin-coater to the hot plate and left there for approximately one hour [23].

### 3.1.5 Hole transporting layer (HTL)

The spiro-OMeTAD HTL is deposited on top of the perovskite layer by following the procedures described by Saliba et al. [23]. The layer is deposited by spin-coating, using a spiro-OMeTAD solution containing two additives. These additives are bis-(trifluoromethane)sulfonimide lithium salt (LiTFSI) and 4-*tert*-butylpyridine (tBP). The quantities and concentrations of the chemicals used to prepare the spiro-OMeTAD solution are obtained from Ossila [60]. Note that Saliba et al. [23] also use tris(2-(1H-pyrazol-1-yl)-4-*tert*-butylpyridine)cobalt(III) tri[bis(trifluoromethane)sulfonimide] (FK209) as an additive. However, because FK209 was unavailable at the time of perovskite solar cell fabrication, the spiro-OMeTAD layers of the cells studied in this work were fabricated without this additive. This will most likely have a negligible impact on the results obtained in this work, as the present study primarily examines the perovskite layer and not the HTL. Other measurements, like I-V curve tracings, would have been more likely to be affected by this since FK209 is a p-type dopant that increases the HTL's conductivity [61].

To prepare 1 mL of the spiro-OMeTAD solution with additives, 85 mg of spiro-OMeTAD is first dissolved in 1 mL of chlorobenzene. The solution is shaken for a few minutes at room temperature. Afterward, 20  $\mu\text{L}$  of a 500 mg/mL LiTFSI in acetonitrile solution and 34  $\mu\text{L}$  of 98% (volumetric percentage) tBP are added [60].

The spin-coating is performed by dropping 50  $\mu\text{L}$  of the spiro-OMeTAD solution with additives in the middle of the substrates while they rotate at 4000 rpm. The substrates are then left spinning for 10 s. The samples are stored in the glovebox until the night before HSPL imaging. This last night, they are

stored in the dark outside of the glovebox to promote lithium doping of the spiro-OMeTAD layer [23].

### 3.2 Hyperspectral Photoluminescence (HSPL) Imaging

This section describes the HSPL imaging setup and its instruments. It also summarizes the process of acquiring HSPL images of the samples.

#### 3.2.1 Instruments

##### Laser

A 444.8 nm line laser is used as the excitation source for the HSPL imaging. The laser model is ILS-450-1000-S-A-20-PF3000-SDP, and it is manufactured by Osela. It generates a line of approximately uniform intensity. The rated output power is 750 mW, but the optical power is adjustable. The fan angle is 20°.

##### Camera

The imaging setup employs a hyperspectral line scan camera, VIS PFD V10E, from Specim. This camera is equipped with a Si-detector. It records 800 wavelength bands ranging from 391.4 nm to 1023.1 nm. The widths of the wavelength bands vary between 0.75 nm and 0.83 nm.

Each sensor on the camera's detector array has 12 bits available for storing its intensity values. Hence, the camera outputs intensity values ranging from 0 to 4095. These values are called *counts*, and they are proportional to the number of photons hitting each sensor [21].

##### Longpass filter

The hyperspectral camera is equipped with a longpass filter to prevent reflected laser light from entering the camera and causing saturation. The filter is important since the intention is to analyze PL signals and not reflected laser signals. Therefore, a longpass filter that blocks all wavelengths shorter than 525 nm and lets the longer wavelengths pass through is mounted onto the camera.

##### Solar simulator

The Sun 2000 solar simulator from ABET Technologies is utilized for illuminating the samples before the HSPL images are recorded. It has a 550 W

## 3.2 Hyperspectral Photoluminescence (HSPL) Imaging

---

mercury-based arc lamp and an AM1.5G filter. The AM1.5G spectrum is the reference solar spectrum used for defining the *standard test conditions* (STC) for PV modules [8, p. 43].

### 3.2.2 Imaging setup

Figure 3.1 shows a picture of the HSPL imaging setup. The camera (A) and the laser (C) are firmly mounted to a rig, and therefore stationary during the image acquisition process. The samples lie on a sample holder (D) that is mounted onto a horizontal translation stage (E).

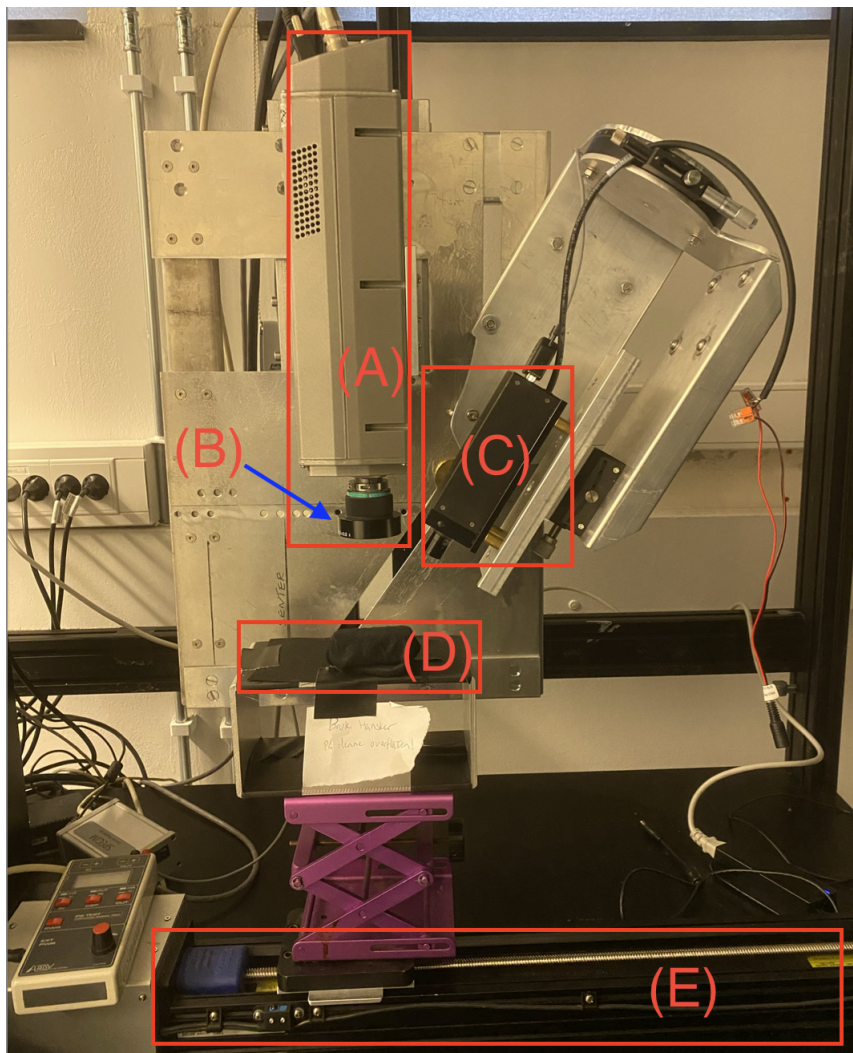


Figure 3.1: *Picture of the HSPL imaging setup. The hyperspectral camera (A), which is equipped with a longpass filter (B), points together with the laser (C) down on the sample holder (D). The sample holder is mounted onto a horizontal translation stage (E). The sample itself is not shown in this image.*



### 3.2.3 Image acquisition procedures

#### Spatial images

Before the hyperspectral images are captured, the camera's focus is adjusted manually to suit the distance between the camera and the sample. Moreover, the speed of the translation stage is adjusted so that the resolutions in the two spatial dimensions become as similar as possible. The desired speed is estimated based on the camera's spatial resolution and the selected frame rate. By studying a test image of a 2.5 cm  $\times$  2.5 cm perovskite sample, it was estimated that the resolution in the detector array's spatial dimension ( $X$ ) is approximately 0.05 mm in the present imaging setup. The frame rate is set to 20 Hz, which means that the camera scans 20 lines per second. Hence, for the second spatial dimension ( $Y$ ) to have the same resolution as the first one, the speed of the translation stage must be 1 mm/s.

Right before the samples are imaged, they are illuminated by the solar simulator for approximately 10 minutes with an irradiance of roughly 1000 W/m<sup>2</sup>. The rationale behind this is to simulate light soaking and induce phase segregation in the samples. Hoke et al. [37] reported that at room temperature, it is sufficient to illuminate MAPb(Br <sub>$x$</sub> I <sub>$1-x$</sub> )<sub>3</sub> ( $0.2 < x < 1$ ) thin films with intensities of less than 1000 W/m<sup>2</sup> for less than a minute to cause redshifts in their PL spectra. These redshifts are associated with light-induced phase segregation. Hoke et al. [37] observed that a high-intensity, lower-energy PL peak formed in less than a minute when using a 457 nm argon laser with an intensity of 150 W/m<sup>2</sup> as the light source. They also found that this spectral change was independent of the spectrum and coherence of the light source [37]. It is hence assumed that using the solar simulator to illuminate the samples with an intensity close to 1000 W/m<sup>2</sup> for a few minutes before the HSPL images are recorded should be sufficient to induce phase segregation in the samples.

When the image is recorded, the output power of the laser is set to its maximum value. Moreover, the shutter is closed for the last 3000 ms of the recording to facilitate applying a background noise correction scheme (see Section 4.2.1).

#### Time series images

The evolution of the PL spectrum from a single line on the sample under light soaking can be studied by recording a HSPL image while the translation stage is held stationary. Such a single-line time series image may give information about the temporal aspect of light-induced phase segregation, by showing how the line reacts to a continuous laser illumination. In this work, single-line time

## 3.2 Hyperspectral Photoluminescence (HSPL) Imaging

---

series images were taken by capturing 800 frames at a 10 Hz frame rate. The exposure time was set to 50 ms. The resulting hypercube has a spatial ( $X$ ), a temporal ( $T$ ), and a spectral dimension ( $\lambda$ ). Light soaking with the solar simulator is not performed before the time series images are recorded so that the sample is in an unsegregated state at the start of the time series recording.

### 3.2.4 Overview of samples and hyperspectral images studied in this work

Table 1 gives an overview of the samples imaged in this work. Two of the samples are full-stack cells, which means that they comprise all layers described in Section 3.1. The last cell, which is a half-stack cell, was fabricated without the HTL. The perovskite layers of samples 1 and 2 were deposited only four days before they were imaged. Moreover, sample 1's HTL was deposited the day before HSPL imaging. These samples are hence referred to as being "fresh". Sample 3's HTL was deposited five days before HSPL imaging on an older perovskite layer. Therefore, this sample's perovskite layer was significantly older than the perovskite layers of the other two samples at the time of HSPL image acquisition. Sample 3 is therefore labeled "old".

Table 1: *Description of the PSC samples imaged in this work. One of the cells was fabricated without the HTL (half-stack). The samples labeled as "fresh" were fabricated only a few days before hyperspectral image acquisition. Sample 3, labeled as "old", was older than the other two samples at the time of HSPL imaging.*

Sample number	Sample type	Substrate	State
1	Full-stack PSC	ITO	Fresh
2	Half-stack PSC	ITO	Fresh
3	Full-stack PSC	ITO	Old

Samples 1 and 3 in Table 1 were used to record spatial images. Sample 2 was used for studying the evolution of single-line PL spectra. Table 2 summarizes the frame rates, exposure times, and translation stage speeds used when recording the HSPL images of the samples.

### 3.3 X-ray Fluorescence (XRF) Imaging

---

Table 2: *Frame rates, exposure times, and translation stage speeds for the HSPL images studied in this work. The table also summarizes the light soaking procedures used before recording the spatial images. The PSC samples are described in Table 1*

Sample	Image type	Light soaking	Frame rate	Exposure time	Speed
1	Spatial	approx. 1000 W/m <sup>2</sup> , 10 min	20 Hz	20 ms	1.0 mm/s
2	Time series	-	10 Hz	50 ms	0 mm/s
3	Spatial	approx. 1000 W/m <sup>2</sup> , 10 min	20 Hz	30 ms	1.0 mm/s

### 3.3 X-ray Fluorescence (XRF) Imaging

XRF images of sample 1 (fresh full-stack PSC) are taken in an attempt to validate the analysis of the HSPL images. The XRF images are taken after the HSPL images on the same day, to ensure that the sample changes as little as possible in the meantime.

The XRF images are recorded using the M4 Tornado  $\mu$ -XRF from Bruker. The instrument uses a X-ray tube with rhodium (Rh) as the target material for generating its X-rays. The  $\mu$ -XRF images a smaller part of the sample. The instrument's spatial resolution is 26  $\mu$ m in the present setup, which gives the images a width of 1.39 cm and a height 1.04 cm. Six cycles are used, which means that the software calculates the results based on six scans of the sample. The dwell time is 30 ms, which means that X-ray signals from a pixel are collected for 30 ms before the detector moves on to the next pixel. A summary of all the settings are given in Figure D.1 in Appendix D.

Different X-ray fluorescence emission lines can be utilized for identifying the various elements [50]. For creating elemental maps of bromide, lead, and iodide in this work, the counts of signals from the so-called  $K_\alpha$ ,  $L_\alpha$ , and  $L_\beta$  lines are used, respectively. These lines represent different electronic transitions between the orbitals of the atoms [50].

---

## 4 Data Processing

This chapter covers the steps involved in analyzing the HSPL images. The Python packages used in this work are listed in Section 4.1. Section 4.2 explains and motivates the preprocessing steps. This includes the background noise correction scheme, conversion from wavelength to energy units, and a normalization scheme for scaling the intensity values. Section 4.3 describes two data analysis pipelines that combine preprocessing steps, PCA, and MCR to produce various outputs. These pipelines were developed and tested on a hyperspectral EL image before the fabrication of perovskite solar cells in the laboratories at NMBU was finished. The hyperspectral EL image was taken of a 2 cm × 2 cm perovskite solar cell of unknown chemistry produced by *Great Cell*. The hyperspectral image comprises 31 wavelength bands of uniform width, spanning from 580 nm to 880 nm. Some results obtained by applying the pipelines to the EL image are given in Appendix C.

### 4.1 Software

The following Python packages were used for analyzing the hyperspectral luminescence images in this work:

- *scikit-learn*, version 1.2.2
- *pyMCR*, version 0.3.2

Some additional packages were used to process and visualize the hyperspectral images:

- *NumPy*, version 1.26.4
- *Matplotlib*, version 3.8.0
- *SciPy*, version 1.11.4

### 4.2 Preprocessing

The preprocessing of the hyperspectral images includes the following steps:

- Reading the image files.
- Background noise correction (only PL images).
- Rotation and cropping.

- Discarding unnecessary wavelength bands (only PL images). The first 250 and the last 150 wavelength bands are discarded to reduce the image processing steps' memory usage and computational cost. This corresponds to keeping only the wavelength bands between 582 nm and 900 nm for analysis.
- Removing negative values that appear due to the background correction scheme.
- Conversion from wavelength (nm) to energy (eV) units.
- Scaling the intensity values by normalizing the areas underneath the pixel spectra.

Python code used to read the image files, rotate and crop the images, and remove negative values is provided in Appendices A.1 and A.2. The rest of this subsection will focus on the background noise correction, the conversion from wavelength to energy units, and the scaling of intensity values.

### 4.2.1 Background noise correction

Before any further processing of the HSPL images can be performed, a background correction scheme is applied to take detector variations into account and remove this background noise from the pixels. The background noise level for each pixel  $(x_i, \lambda_k)$  on the camera's detector array is calculated and accounted for by applying an algorithm developed by Flø [62] and described by Mehl [21].

First, when the HSPL image is acquired, the last lines that the line scan camera records are captured with the shutter closed. These dark frames contain no PL signals and can thus be used to calculate the pixels' background noise levels [21]. In this work, the shutter was closed for the last three seconds of each recording (see Section 3.2.3).

Second, the background noise levels for each detector pixel are calculated and stored in a background noise matrix,  $BNM$ . The background noise levels are computed by taking the median of the last 50 dark frames in the hyperspectral image. The entry in the  $BNM$  corresponding to the detector pixel  $(x_i, \lambda_k)$  is therefore calculated by taking the median of the values from  $(x_i, y_j, \lambda_k)$  to  $(x_i, y_{j+50}, \lambda_k)$  in the hypercube [21].

Third, the  $BNM$  is subtracted from each raw frame  $RF(Y)$  in the hyperspectral image to yield corrected frames  $CF(Y)$  [21],

$$CF(Y) = RF(Y) - BNM \quad (4.1)$$

Python code used to perform the background noise correction can be found in Appendix A.2.1.

#### 4.2.2 Jacobian conversion from wavelength (nm) to energy (eV) units

Raw hyperspectral luminescence images contain intensity values measured within different wavelength bands (in nm). As Mooney and Kambhampati [63] state, the recorded signals are intensity values per unit wavelength. They argue that this is important to consider when converting from a wavelength to an energy scale since evenly spaced intervals in the wavelength domain become unevenly spaced in the energy domain. The reason is the inverse relationship between wavelength and energy. Without transforming the signals properly, the total luminescence signal obtained by integrating a spectrum in the wavelength domain will differ from the value found by integrating the spectrum in the energy domain. This may lead to incorrect results [63].

The wavelength scale is converted into an energy scale by applying the following equation:

$$E = \frac{hc}{\lambda}. \quad (4.2)$$

Here,  $E$  is the energy,  $h$  is Planck's constant,  $c$  is the speed of light, and  $\lambda$  is the wavelength [63].

The recorded signals,  $f(\lambda)$ , are transformed into signals  $f(E)$  in the energy domain by applying the following equation:

$$f(E) = -f(\lambda) \cdot \frac{hc}{E^2}. \quad (4.3)$$

This conversion is referred to as the *Jacobian transformation*. It is derived by differentiating Equation (4.2) and imposing equal integrated (total) signals of  $f(\lambda)$  and  $f(E)$  [63]. A Python implementation of the Jacobian transformation can be found in Appendix A.3.

### 4.2.3 Scaling the intensity values by normalizing the areas underneath the pixel spectra

EL/PL intensity values may differ significantly among the pixels in a raw hyperspectral image. These variations may be attributed to factors like variations in film thickness and material defects, as discussed in Section 2.4.1. Phase segregation will also influence the EL/PL intensities, as it increases the non-radiative recombination rates by forming trap states within the bandgap [4]. Moreover, the iodide-rich domains that form during phase segregation typically emit strong luminescence signals [37].

When the aim is to use PCA for mapping phase segregation in a PSC, it may be problematic if variations in the EL/PL intensity dominate the total variance in the image. The reason is that it would require a more comprehensive analysis to quantify the individual contributions from factors like film thickness and phase segregation to these EL/PL intensity variations. Principal components are constructed to explain as much of the variance in the image as possible, and if EL/PL intensity variations dominate the total variance in the image, the principal components will capture these variations. The problem is that the intensity variations captured by the principal components can be caused by other factors than phase segregation.

Luckily, as was mentioned in Section 2.2.3, phase segregation also leads to local bandgap variations, which in turn cause shifts in the EL/PL spectra. Consequently, it is believed that shifts in the EL/PL spectra are even more relevant than mere intensity changes for mapping phase segregation with PCA. Therefore, to ensure that EL/PL intensity variations caused, for instance, by variations in film thickness do not obscure this important information, it may be beneficial to scale the intensity values before performing PCA.

One approach is to normalize the areas underneath the pixel spectra. This means that all pixels will have equal total intensities. The rationale behind this is to limit the influence of factors like film thickness while at the same time preserving the shapes of the spectra. The procedure includes the following three steps:

**1) For every pixel, calculate the total luminescence intensity by applying the trapezoidal rule for numerical integration.** This numerical integration scheme uses straight line segments to approximate a curve between a set of discrete partition points [64, p. 504]. The area  $A$  underneath a curve is calculated numerically by applying the equation

$$A = \frac{\Delta\lambda}{2}(y_0 + 2y_1 + 2y_2 + 2y_3 \dots + 2y_{n-2} + 2y_{n-1} + y_n). \quad (4.4)$$

Here,  $\Delta\lambda$  represents the mesh size, which is the distance between two successive partition points along the wavelength axis. The values  $y_0, y_1 \dots y_n$  represent the given spectrum's intensity values at the partition points [64, p. 504].

If the hyperspectral image comprises uniformly spaced wavelength bands, Equation (4.4) can be applied directly to the pixel spectra to integrate them. In this situation,  $\Delta\lambda$  is simply the distance between two successive wavelength bands in the hyperspectral image. However, if the spacing between the wavelength bands is non-uniform, a uniform subdivision of the interval is required before applying Equation (4.4). This is achieved by applying piecewise linear interpolation between the hyperspectral image's wavelength bands, which means that the spectra are approximated by straight line segments between their mesh points [65, p. 7].

**2) To normalize the spectra, each spectrum is divided by its area A.** Mathematically, this can be expressed as

$$f_{norm}(\lambda) = \frac{1}{A}f(\lambda). \quad (4.5)$$

Here,  $f(\lambda)$  is the original spectrum, and  $f_{norm}(\lambda)$  is the normalized spectrum. The effect of Equation (4.5) is that the area underneath every scaled spectrum will be 1.

**3) Finally, the normalized spectra  $f_{norm}(\lambda)$  in the wavelength domain can be converted to the energy domain by applying Equation (4.3).** As discussed earlier, the Jacobian transformation preserves the areas underneath the curves. Hence, the resulting spectra are also normalized.

The Python code used to perform interpolation and numerical integration, as well as the subsequent normalization of the areas underneath the spectra, can be found in Appendix A.4.

When it comes to performing MCR, choosing to scale the intensity values or not before analysis to account for intensity variations among the pixels should not make much of a difference. The reason is that MCR calculates individual weights (concentrations) for the pure components in all pixels. These concentrations let the model scale its reconstructions of the pixel spectra. Therefore, in theory, it should not matter whether the pixel spectra are scaled before



performing MCR or not, since the concentrations act as independent scaling factors anyway. The only difference is regarding the interpretation of the results. The question is whether one wants to take intensity variations into account or not when comparing the concentrations of the MCR components in different pixels.

The EL image used for developing the data analysis techniques had an intensity gradient in the horizontal direction, as can be seen in Figure C.1 in Appendix C. Since layers deposited by blade coating typically show similar thickness gradients in the coating direction [66], it is likely that the PSC was fabricated using blade coating. Hence, it is assumed that the intensity gradient in the EL image is a consequence of thickness variations caused by this deposition method. Normalization of the areas underneath the pixel spectra was therefore performed before PCA and MCR were applied to the EL image.

PCA is performed both with and without normalizing the areas for the HSPL images of the samples produced in REALTEK's laboratories. This is done to examine the effect of the normalization scheme and to check if it enhances the PCA-based analysis. However, MCR is performed without normalizing the spectra.

### 4.3 Data Analysis Pipelines

Figure 4.1 illustrates the steps involved in the two data analysis pipelines. The first and simplest pipeline comprises only a PCA step after the preprocessing steps. The second pipeline uses the PCA loadings to initialize an MCR analysis step. These two pipelines will be described in greater detail in Sections 4.3.1 and 4.3.2. Python code snippets showing how the individual analysis steps are implemented are provided in Appendix B.

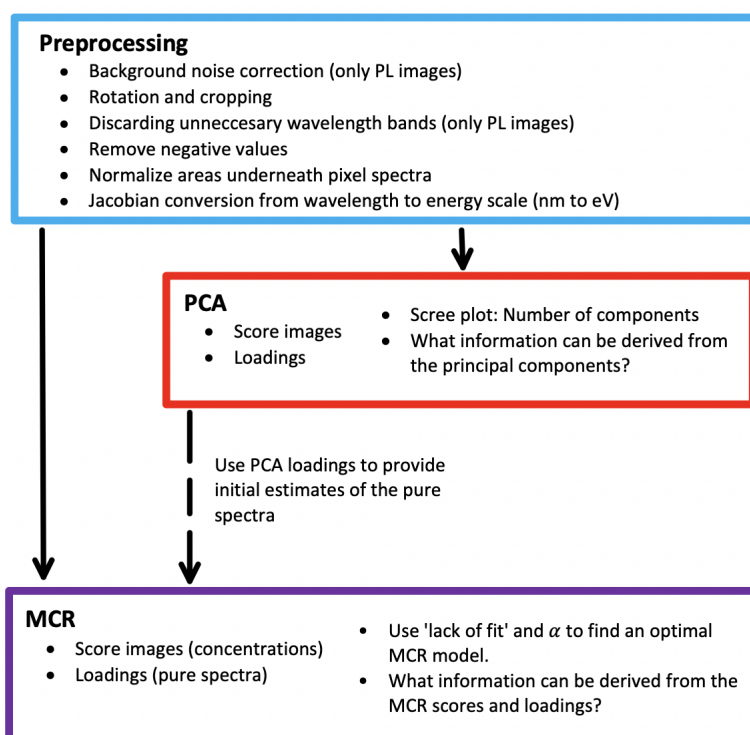


Figure 4.1: A graphical representation of the steps involved in the two data analysis pipelines. The arrows show how data flows between their building blocks. A solid arrow indicates that the output from one step is passed on as input to the next step. The dashed arrow indicates that the PCA loadings are used to generate initial guesses of the pure spectra required in the MCR step. Each text box summarizes the given analysis step, including the output it produces, analytical tools to help select optimal parameter settings, and some key questions that will be answered when examining the outputs from the pipelines.

### 4.3.1 Pipeline 1: PCA

Performing PCA on the preprocessed hyperspectral images produces several types of output that can be used to extract information about the imaged PSC. This includes:

- A *scree plot*, showing the variance explained by each principal component. It can provide an indication of the number of principal components to keep for further analysis.
- The *loadings*, showing the spectral information captured by each principal component.
- The *score images*, showing the spatial information in each principal component.

The variance explained by each principal component is given by the eigenvalues of the covariance matrix associated with the data matrix  $\mathbf{D}$ , as discussed in Section 2.6.1. The *explained variance ratio*,  $EVR$ , of principal component number  $j$  is calculated using the following equation:

$$EVR_j = \frac{\lambda_j}{\sum_{i=1}^d \lambda_i}. \quad (4.6)$$

Here,  $\lambda_i$  represents the eigenvalue associated with principal component number  $i$ . Moreover,  $d$  is the total number of principal components, which equals the number of wavelength bands in the hyperspectral image [53, p. 147, 151].

### 4.3.2 Pipeline 2: PCA + MCR

After applying PCA, the PCA loadings aid the initialization of the MCR algorithm. To initialize the method with  $N$  MCR components, initial guesses of  $N$  pure spectra are provided. In this work, these  $N$  spectra are evenly spaced and Gaussian-shaped. The role of the PCA loadings is to help locate the peaks of these spectra. This can be explained by using the initialization of the MCR algorithm applied to the EL image in Appendix C as an example.

Figure 4.2 (a) shows the PC1-loadings calculated by performing PCA on the normalized hyperspectral EL image. The PC1 line has two extremes, a maximum at 1.59 eV and a minimum at 1.70 eV. From the scree plot in Figure C.2 in Appendix C.2, it is evident that PC1 explains more than 90% of the total variance in the normalized image. Therefore, it is likely that the two aforementioned spectral peaks are important components of the total EL signal

### 4.3 Data Analysis Pipelines

---

from the perovskite material. The leftmost and rightmost spectra among the  $N$  spectral guesses are therefore generated so that their peaks coincide with these PC1-extremes. If MCR is to be initialized with more than two components, the remaining spectra are placed between these two spectra, equidistant from each other. Figure 4.2 (b) shows the resulting initial guesses when using  $N = 5$ . The Python implementation of this procedure is given in Appendix B.2.

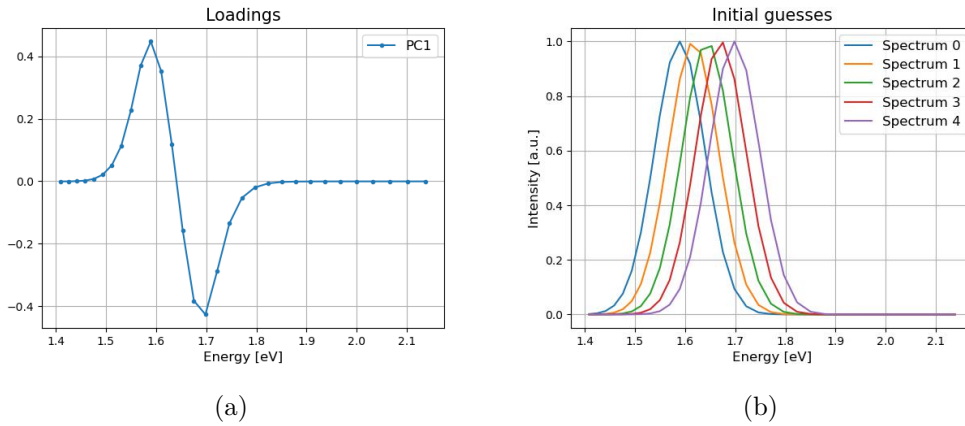


Figure 4.2: *The method used to initialize the MCR algorithm. (a) PC1-loadings obtained by performing PCA on the normalized hyperspectral EL image. (b) Example of spectral guesses used to initialize the MCR algorithm. Here,  $N = 5$  components are used. Spectrum 0 and Spectrum 4 have peaks that coincide with the PC1 extremes shown in (a).*

Running the MCR algorithm produces two types of output:

- *Loadings*, which represent the pure spectra of each component.
- *Score images*, which represent the concentrations of each component in every pixel.

The MCR algorithm will be initialized with different numbers of components to determine an adequate number. Moreover, Ridge Regression is used, and different values of the regularization parameter  $\alpha$  are tested to help discard models that are likely to suffer from overfitting. The following paragraphs introduce two metrics that will help choose the optimal number of components.

### Using the residuals to determine an optimal number of MCR components

An indication of the appropriate number of MCR components can be obtained by using the residuals to calculate the reconstruction error of MCR models initialized with different numbers of components. Jaumot et al. [58] define a "lack of fit" metric to quantify the differences between the original spectra (in  $\mathbf{D}$ ) and the reconstructed spectra (in  $\mathbf{CS}^T$ ). "Lack of fit" ( $LoF$ ), expressed as a percentage, is calculated as follows:

$$LoF(\%) = 100 \sqrt{\frac{\sum_{i,j} e_{ij}^2}{\sum_{i,j} d_{ij}^2}}. \quad (4.7)$$

In this equation,  $d_{ij}$  is the entry in the  $i$ th row and  $j$ th column of the data matrix  $\mathbf{D}$ , and  $e_{ij}$  is the corresponding entry in the residual matrix  $\mathbf{E}$  [58].

For examining the reconstruction errors at the level of individual pixels, a pixel-wise "lack of fit" ( $PLoF$ ) is defined in this work. The "lack of fit" value  $PLoF_i(\%)$  for the pixel with index  $i$  is calculated by using the following equation:

$$PLoF_i(\%) = 100 \sqrt{\frac{\sum_j e_{ij}^2}{\sum_j d_{ij}^2}}. \quad (4.8)$$

Here, the only difference compared to Equation (4.7) is that the summation is performed only along the spectral dimension of the residual matrix  $\mathbf{E}$ , represented by the index  $j$  (the columns). The standard  $LoF(\%)$  involves summation along both the spatial (the rows,  $i$ ) and spectral (the columns,  $j$ ) dimensions.

### Constraints, objective function, and convergence criteria

The *pyMCR* package [59] allows applying different constraints to  $\mathbf{C}$  and  $\mathbf{S}^T$ . In this work, non-negativity constraints are applied to both matrices. This is *pyMCR*'s default settings.

Regarding the convergence/stopping criteria, *pyMCR* uses the  $MSE$  objective function (Equation (2.6)) and a combination of several conditions for terminating the optimization process [59]. The optimization process stops if

- a preset maximum number of iterations is reached. Default: 50 iterations (equivalent to 100 *half-iterations*).
- the error ( $MSE$ ) increases more than a preset factor between two half-iterations. For instance, if the user-defined factor is 1 (100%), the optimization process terminates if the error more than doubles between two half-iterations [59]. Default: 0.0 (i.e. the algorithm terminates if the  $MSE$  increases).
- the error increases for more than a preset number of consecutive iterations. Default: 10 iterations.
- the error changes less than a preset value per iteration. Default: none (the default is not to use this criterion).
- more than a preset number of half-iterations is performed without the algorithm reaching a new error-minimum. Default: 10 half-iterations.

After the algorithm terminates, the matrices  $\mathbf{C}_{[k]}$  and  $\mathbf{S}_{[k]}^T$  associated with the smallest error (in terms of  $MSE$ ) are kept for further analysis.

To prevent the optimization process from terminating whenever the error increases between two half-iterations, a 500% increase in the error was tolerated in this work. *pyMCR*'s default settings were used for the rest of the stopping criteria.

---

## 5 Results and Discussion

The main purpose of this work is to investigate whether HSPL imaging and the accompanying data analysis techniques can map light-induced phase segregation in solar cell samples. To answer this question, a time series image of one of the samples is first analyzed to investigate the evolution of the PL signals under light soaking. Thereafter, spatial images are analyzed to study the PL signals from the entire sample's surface. These images are analyzed using PCA and MCR. An integral part of the work is to build an MCR model that seems capable of capturing relevant information from the HSPL images. The PL signals detected by the chosen MCR model are then compared to bandgaps of different halide stoichiometries as predicted by the regression model in Equations (2.1) and (2.2). This is done in an attempt to estimate which segregated phases that may be present in the samples. The results from the XRF imaging are also presented, and their potential as a means of confirming phase segregation is discussed. The results from this work are discussed and compared to similar findings in other studies.

### 5.1 Estimating the Perovskite Samples' Bandgaps

By using the linear regression model from Li et al. [32] in Equation (2.1) and the stoichiometric coefficients presented in Section 3.1.3, the perovskite's bandgap is estimated to be 1.60 eV. This is calculated using a bromide-to-iodide ratio of  $y = 0.15$ , which corresponds to the sample's overall ratio between these two halides. The model in Equation (2.1) could also be used to predict the perovskite's bandgap for other halide compositions. Figure 5.1 shows the bandgap  $E_G$  calculated by using Equations (2.1) and (2.2) as a function of the bromide-to-iodide ratio  $y$ . All other stoichiometric coefficients are kept constant. The relationship is not exactly linear because of the factor  $R$  in Equation (2.2), but as can be seen in the figure, the relationship between  $y$  and  $E_G$  is very close to linear in the relevant domain. For a sample with only iodide ( $y = 0$ ), the model estimates the bandgap to be 1.50 eV (corresponding to 827 nm). A sample containing only bromide ( $y = 1$ ) on the other hand, would have a bandgap at 2.17 eV (571 nm).

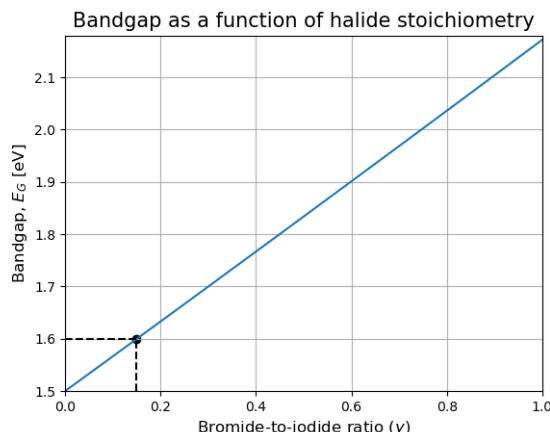


Figure 5.1: *The bandgap  $E_G$  as a function of the bromide-to-iodide ratio  $y$ . The bandgap is calculated using the model from Li et al. [32] (Equations (2.1) and (2.2)). The black dot shows the sample’s overall bromide-to-iodide ratio (0.15) and the corresponding estimated bandgap (1.60 eV).*

## 5.2 Time Series of Single-Line PL Spectra (Sample 2)

Figure 5.2 shows PL spectra at selected timestamps obtained by recording a time series image of a single line on sample 2. Each spectrum is calculated as the mean of the line’s pixel spectra (i.e. by taking the mean in the  $X$  direction) at a given timestamp  $T$ . The spectrum representing the recording’s first frame is given the timestamp  $T = 0$  s. The last frame has the timestamp  $T = 80$  s because the recording consists of 800 frames recorded at a 10 Hz frame rate. The red arrow in the figure points from  $T = 0$  to  $T > 0$ . Note that the spectra in Figure 5.2 are calculated based on raw intensity values (i.e. without performing the background correction). This gives them a vertical offset caused by background noise from the detector array.

The spectrally accumulated time series image of sample 2 is shown in Figure E.1 in Appendix E. It is calculated by summing the counts in all wavelength bands for every pixel  $(X, T)$  in the image.



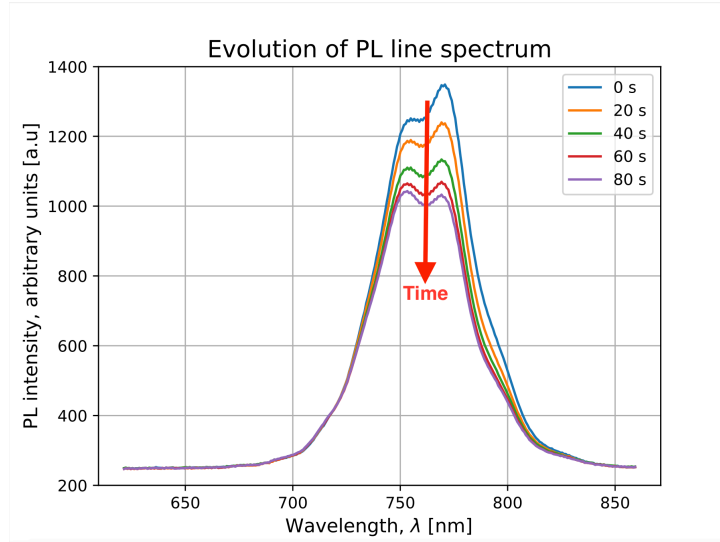


Figure 5.2: *PL line spectra calculated by taking the mean of the line’s pixel spectra at different timestamps  $T$ .  $T = 0$  s represents the recording’s first frame. The total duration of the recording is 80 seconds.*

Figure 5.2 shows that the line’s PL intensity decreases as time passes. The PL intensity decreases at a higher rate at the start of the recording than at the end. Moreover, one can observe that the spectrum’s shape changes from the first to the last frame. In the first frame ( $T = 0$  s), the line’s PL maximum is located at around 770 nm (1.61 eV), which is close to the sample’s estimated bandgap from Section 5.1 (1.60 eV). The spectrum also has a local maximum located at a shorter wavelength (approximately 755 nm, corresponding to 1.64 eV). As time passes, the shorter-wavelength peak grows relative to the longer-wavelength peak. Consequently, the line’s PL maximum is located at 753 nm (1.65 eV) at the last timestamp ( $T = 80$  s). The line’s PL spectrum hence undergoes a blueshift between the recording’s first and last timestamps.

### Discussion: Evolution of the line’s PL spectrum

Hoke et al. [37] demonstrated that the total PL spectra of mixed-halide perovskite solar cells typically redshift under light soaking. They observed that a lower-energy peak formed in the PL spectra of  $\text{MAPb}(\text{Br}_x\text{I}_{1-x})_3$  perovskites with  $0.2 < x < 1$  under light soaking (note that  $x$  here is the same as  $y$  in Equation (2.1)). This peak dominated the original PL peak after less than a minute of light soaking. Moreover, they found that the overall PL intensity increased greatly during light soaking. Their explanation of these phenomena was that small lower-bandgap, iodide-rich domains with high PL efficiencies formed upon light exposure [37].

The PL spectra in Figure 5.2 represent only one line on the sample, and not the entire perovskite surface as the PL spectra from Hoke et al. [37] do. Therefore, the blueshift and decrease in PL intensity shown in Figure 5.2 do not necessarily contradict the results from Hoke et al. [37]. One explanation may be that iodide ions migrate away from the imaged line and form iodide-rich domains in other sample regions. This outward flux of iodide must be accompanied by an inward flux of bromide to conserve the stoichiometry of the perovskite structure [17]. Consequently, the bromide-to-iodide ratio in the line will increase, resulting in a higher bandgap (see Section 2.2.2) and a corresponding blueshift of the line's PL spectrum.

Even though it is difficult to confirm phase segregation based on only one line on the sample, it is clear from Figure 5.2 that light soaking influences the line's PL spectrum. It is, therefore, possible that the perovskite samples studied in this work undergo phase segregation under light exposure, but it is unknown how pronounced the effect is. Other studies have shown that phase segregation mainly occurs in perovskites with large bromide fractions [4]. The perovskite studied in this work has a bromide-to-iodide ratio of 0.15, which is for instance lower than the 0.2 to 1 range where Hoke et al. [37] observed phase segregation in their  $\text{MAPb}(\text{Br}_x\text{I}_{1-x})_3$  perovskite samples. The perovskite samples studied in this work may therefore be less prone to light-induced phase segregation than other mixed-halide perovskites, making the effect of phase segregation less pronounced than in other samples.

Additionally, since the rate at which the PL intensity decreased in Figure 5.2 was lower at the end of the recording than at the start, it could be that the perovskite material is about to stabilize and reach an equilibrium state not too long after the 80 s recording ended. If so, the 10-minute light soaking step used before recording the spatial images in this work should be sufficient for inducing phase segregation in the samples, if it occurs.

## 5.3 Raw Spatial HSPL Image (Sample 1)

### 5.3.1 Accumulated PL intensities

Figure 5.3 shows the accumulated PL intensity in each pixel of the spatial hyperspectral image of sample 1. The values are obtained by summing the counts in all wavelength bands. This image shows that there is a certain signal from the background pixels in the image.

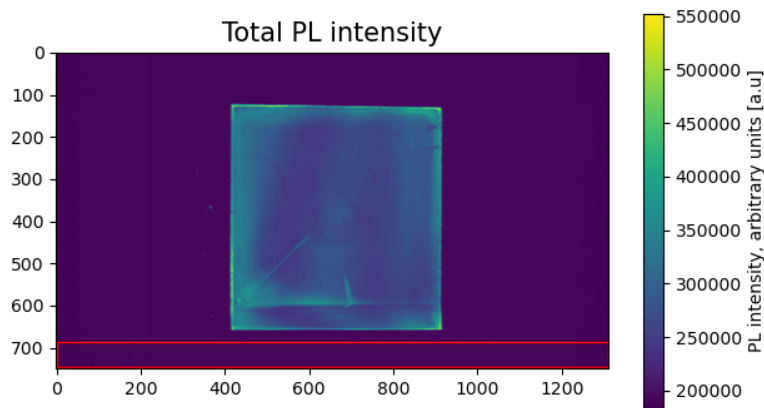


Figure 5.3: *Total PL intensity map of sample 1. The total intensity is calculated pixel-wise by summing the counts in all wavelength bands. The red rectangle marks the dark frames at the end of the recording.*

Something unexpected happens around line 600 in the image in Figure 5.3, where a horizontal line stretches across the entire sample surface. Additionally, the orientation of the sample's outer edge suddenly changes. The reason why this happens is not completely understood. Since the region below line 600 looks like a compressed, mirrored version of the region above line 600 (see also Figure 5.5), one explanation may be that the translation stage turned around too early and started recording the sample in the backward direction. This was not noticed when the image was recorded, and the cause for this behavior has not been identified. However, before analyzing the image using PCA and MCR, it is decided to crop the image so that the lower part of the image is removed. The remaining sample area is close to quadratic, so the camera should have covered almost the entire sample surface before the translation

stage turned around.

### 5.3.2 Mean PL spectrum

Figure 5.4 shows the mean spectrum of the entire sample 1 image, calculated by averaging all pixel spectra in the image. The figure also includes dashed lines showing the wavelength of the laser and the cut-on wavelength of the longpass filter.

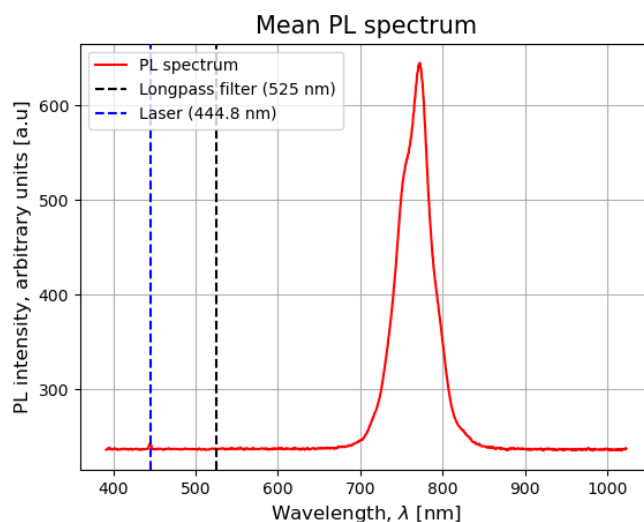


Figure 5.4: Mean PL spectrum of the hyperspectral image of sample 1. The blue, dashed line represents the laser's wavelength (444.8 nm). A local intensity maximum is observed at this wavelength. The black, dashed line represents the cut-on wavelength of the longpass filter (525 nm).

The spectrum's global PL intensity maximum is located at 772.44 nm. This wavelength band is plotted in Figure 5.5. This figure suggests that the lower part of the sample area is a mirrored version of the upper part, as discussed in the previous subsection. Also, a dark circle can be observed in the middle of the substrate. This circle is most likely caused by the dropping of antisolvent during the spin-coating deposition of the perovskite layer. Saliba et al. [23] reported that a cavity can be created in the middle of the substrate if the antisolvent is dropped too rapidly onto the substrate or if it is not dropped in the middle of the substrate.

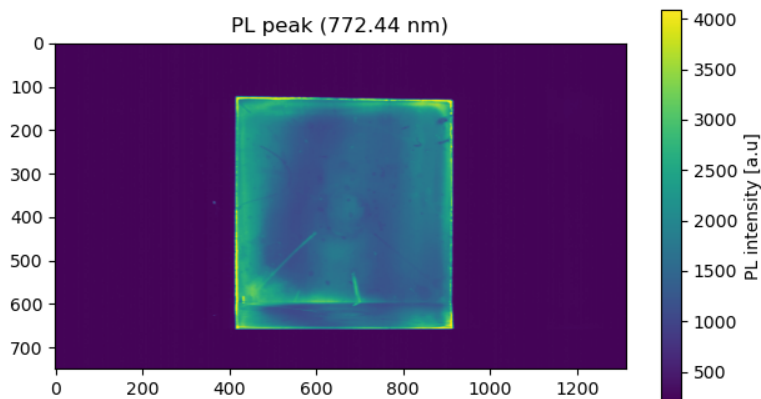


Figure 5.5: *The wavelength band corresponding to the PL maximum of the sample 1 image (772.44 nm). The lower part of the sample area appears to be a compressed, mirrored version of the rest of the sample area.*

Another peak in the PL spectrum in Figure 5.4 is located at the laser’s wavelength. This means that the longpass filter does not completely block the laser light. Luckily, the laser light does not seem to disturb the rest of the signal. The second-order diffraction maximum of the laser light’s wavelength (444.8 nm) overlaps with the first-order diffraction maximum of light with twice the wavelength [22], namely 889.6 nm. Transmitted laser light could, therefore, potentially have disturbed the signal at this wavelength. However, this effect is not observed in the PL spectra.

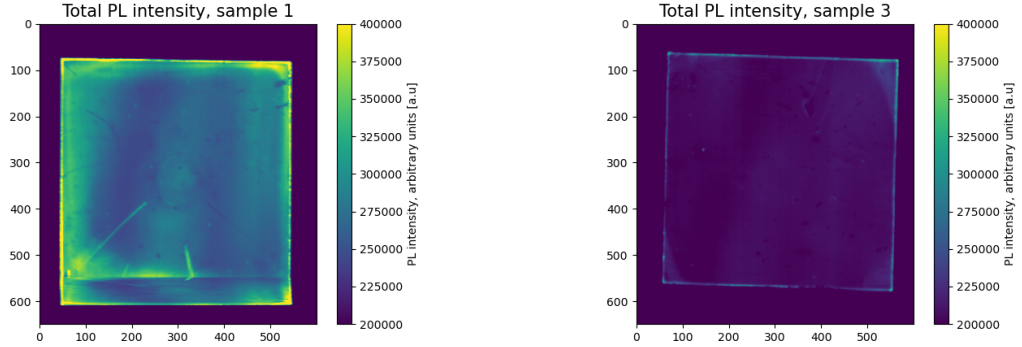
Furthermore, the signal has a vertical offset at around 240 counts, caused by background noise. This background signal can also be seen in the areas outside of the sample in Figure 5.5.

#### 5.3.3 Comparing PL intensities from sample 1 and sample 3

Figure 5.6 shows total PL intensities and mean PL spectra calculated based on the raw spatial images of samples 1 and 3. Figures (a) and (b) show that the PL signal from sample 1 (the fresh sample) is much stronger than the signal from sample 3 (the older sample), even though the exposure time was larger in the sample 3 image (30 ms vs. 20 ms). The mean PL spectra in (c) and (d) are normalized with respect to the exposure time. By comparing these spectra,

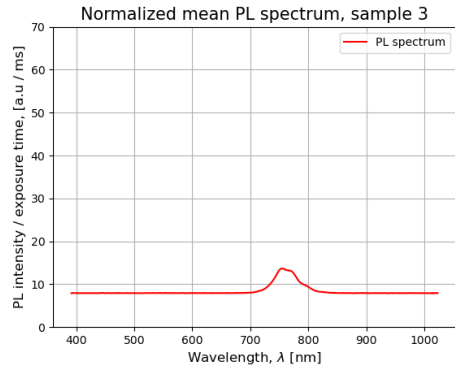
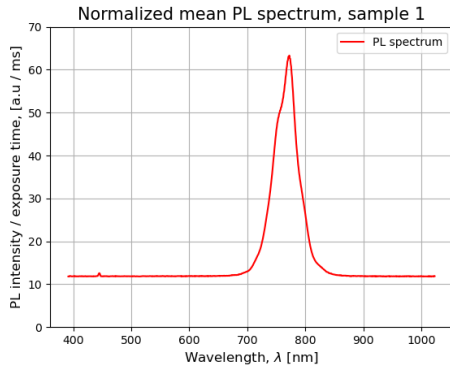
### 5.3 Raw Spatial HSPL Image (Sample 1)

it is evident that the PL signal is much stronger relative to the background signal in (c) than in (d).



(a) Total PL intensity (sample 1)

(b) Total PL intensity (sample 3)



(c) Norm. mean PL spectrum (sample 1)

(d) Norm. mean PL spectrum (sample 3)

Figure 5.6: *Total PL intensities (obtained by summing the counts in all wavelength bands) and normalized mean PL spectra calculated based on images of samples 1 and 3. The spectra in (c) and (d) are normalized with respect to the exposure time. Note that the mean spectra are calculated by averaging all pixel spectra in the windows plotted in (a) and (b), instead of averaging over the entire image as was done in Figure 5.4.*

One explanation for these differences is that sample 3's perovskite layer has had more time to degrade. The sample was stored in a nitrogen-filled glovebox most of the time between device fabrication and HSPL imaging, but degradation may still have occurred. Another possible explanation is that the quality of sample 3, for some reason, was poorer than that of sample 1 after fabrication. Many of the samples fabricated in this work gave PL signals comparable to those from sample 3, and only a few samples gave as strong signals as sample 1. Differences induced by the fabrication process may, therefore, play a role in what is observed in Figure 5.6. Since the PL signal is strongest in the image

of sample 1, this image is used for testing the data analysis techniques in the coming subchapters.

### 5.4 XRF Images (Sample 1)

Figure 5.7 shows the elemental maps obtained by performing XRF imaging of sample 1 using the procedures described in Section 3.3, along with the image captured by the video camera inside the chamber. In the video camera image in Figure 5.7 (a), one can observe the circular region in the middle of the substrate that was discussed in Section 5.3.2. Pinholes are also observed in this image, and they can also be observed to a greater or lesser extent in the elemental maps in Figures 5.7 (b)-(d). In addition to these details, it looks like there is a region with a lower iodide concentration to the left of the circle in the center of the substrate in Figure 5.7 (d). This region is marked with a yellow rectangle.

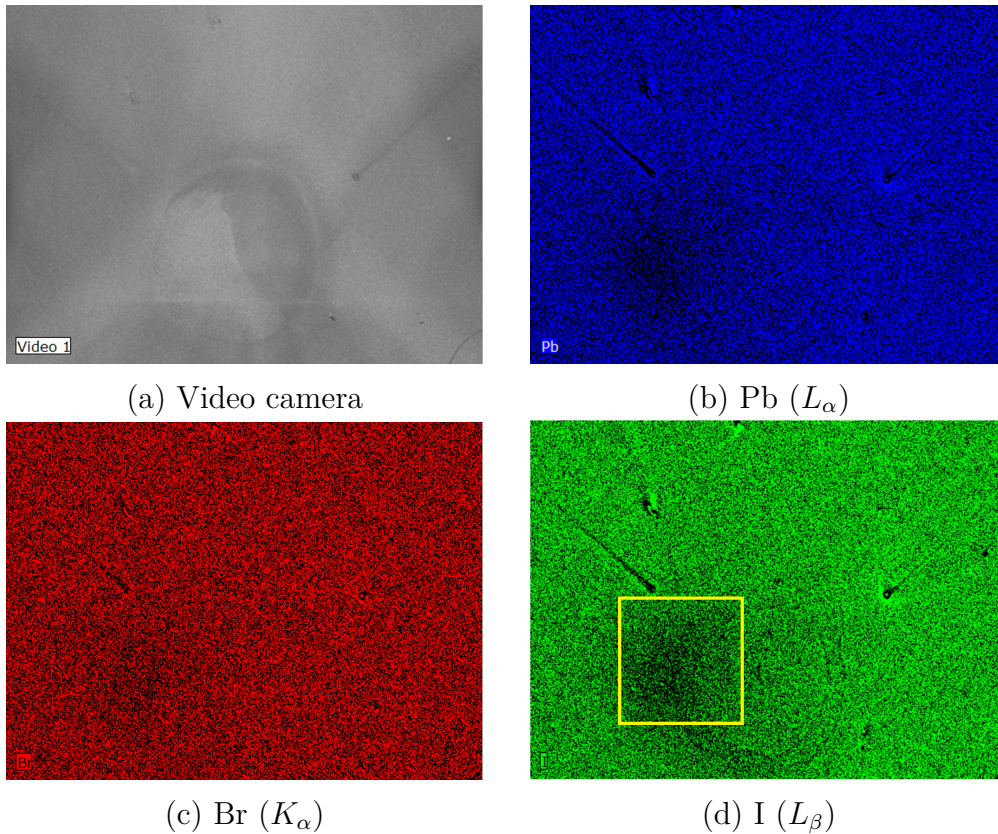
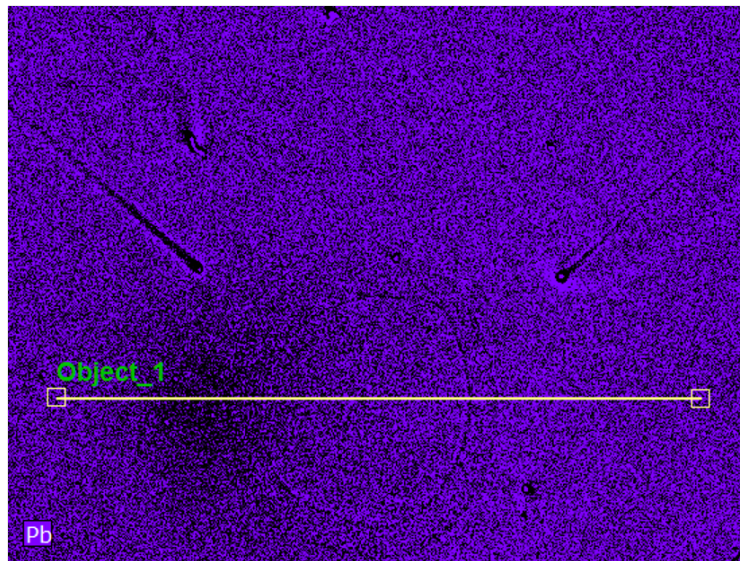


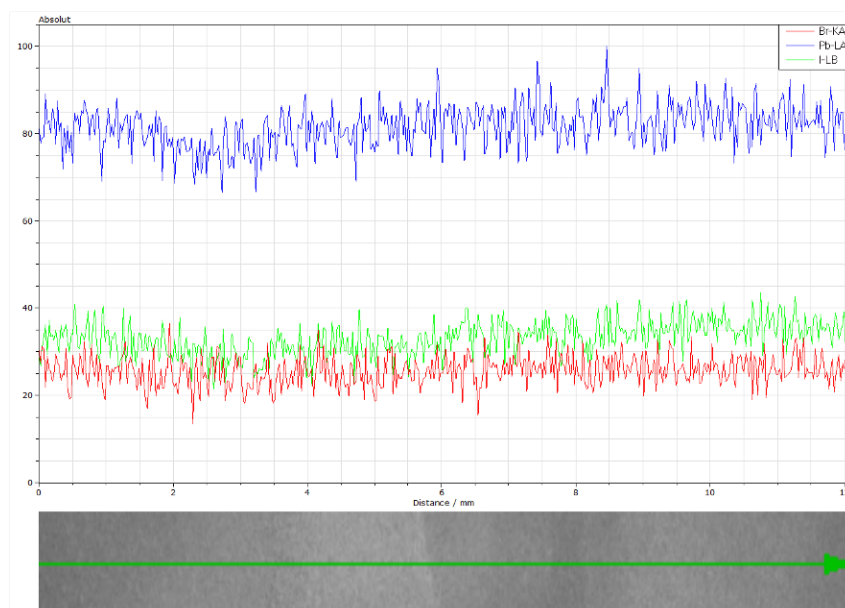
Figure 5.7: *Elemental maps obtained by performing XRF imaging of sample 1. Figure (a) shows the image captured by the video camera inside the chamber. Figures (b)-(d) show the elemental maps of lead, bromide, and iodide. A brighter color indicates a higher concentration. The yellow rectangle in Figure (d) marks a region where the iodide concentration potentially is lower.  $L_\alpha$ ,  $K_\alpha$ , and  $L_\beta$  are the XRF emission lines used to produce the various elemental maps (see Section 3.3). Each image is  $1.39\text{ cm} \times 1.04\text{ cm}$  (i.e. they cover roughly a quarter of the  $2.5\text{ cm} \times 2.5\text{ cm}$  substrates). Note that the sample is imaged upside down compared to the HSPL images in Section 5.3.*

The magnitude of the variations in counts is examined by drawing a line between two points in the image and plotting the counts as a function of position along this line. This is shown in Figure 5.8. The line passes through the previously mentioned area that potentially has a lower iodide concentration.





(a)



(b)

Figure 5.8: Counts in the elemental maps from Figure 5.7 as a function of position along a selected line. Figure (a) shows this line ( $\text{Object}_1$ ) in the Pb map. The counts as a function of position along the line (in mm) are shown in (b). The blue graph represents lead, the green graph represents iodide, and the red graph represents bromide.

In Figure 5.8 (b), it looks like the concentration of lead (measured as X-ray counts) is lower in the aforementioned region, which lies between 2 mm and 3.5 mm away from the point where the line starts. The iodide concentration

also seems to have a similar dip in this region. A lower bromide concentration is not observed in this region. However, the rapid fluctuations of the counts between neighboring pixels make the interpretation of these results challenging. The differences between different sample regions are small compared to these fluctuations. Based on the PCA and MCR analysis of the hyperspectral EL image in Appendices C.2 and C.3, it was expected that one could observe larger sample regions of different bandgaps. It was hypothesized that these bandgap variations were caused by differences in the halide stoichiometry between different sample regions. Such variations in the halide concentrations were not possible to detect using the present XRF setup. Therefore, the XRF images could not be used to confirm systematic variations in the halide concentrations between different macroscopic sample areas. This makes using the XRF images to validate the HSPL image analysis in the upcoming sections difficult.

It also has to be noted that the resolution of the XRF images is low compared to the expected size of the iodide-rich domains based on the literature. As previously mentioned, Bischak et al. [38] estimated the diameter of the iodide-rich clusters in  $\text{MAPb}(\text{Br}_{0.9}\text{I}_{0.1})_3$  to be 8-10 nm. A few years earlier, Hoke et al. [37] estimated that the iodide-rich minority domains comprised crystallites that were at least 50 nm in size. They also claimed that the phase-segregated domains were most likely larger, as each domain may contain more than one crystallite [37]. However, both aforementioned numbers are approximately three orders of magnitude smaller than the spatial resolution of the XRF setup (26  $\mu\text{m}$ ) and the HSPL setup (approx. 50  $\mu\text{m}$ ). This may complicate the detection of the iodide-rich domains if not larger aggregates of sizes comparable to the XRF and HSPL resolutions are present. It may be that a higher resolution setup is needed to be able to study concentration variations on length scales smaller than the pixels in the XRF images in this work. Bischak et al. [38] were, for example, able to map iodide-rich domains near grain boundaries by taking cathodoluminescence images of areas that were approximately 10  $\mu\text{m} \times 10 \mu\text{m}$  in size.

The recorded XRF spectrum and the XRF software's estimated concentrations of the various atomic species in sample 1 are given in Figures D.2 and D.3 in Appendix D. One can observe that the software estimates the concentration of iodide to be more than nine times that of lead. This is clearly wrong, as the combined concentration of the halides should be three times that of lead. Hence, the concentrations estimated by the XRF software must be interpreted with caution.

## 5.5 Preprocessing (Sample 1)

### 5.5.1 Background noise correction

Figure 5.9 demonstrates the effect of the background noise correction scheme described in Section 4.2.1. Figure 5.9 (a) shows the background noise matrix ( $BNM$ ). This matrix shows the background noise for all pixels on the camera's detector array. In Figure 5.9 (b), the histogram corresponding to the  $BNM$  is shown. The mean value in the  $BNM$  is 236, and the standard deviation is 6. This means that most of the detector pixels have noise levels between 230 and 242. However, the minimum and maximum values in the  $BNM$  are 0 and 721.5, respectively. Figures 5.9 (a) and (b) therefore emphasize the need for the background correction scheme.

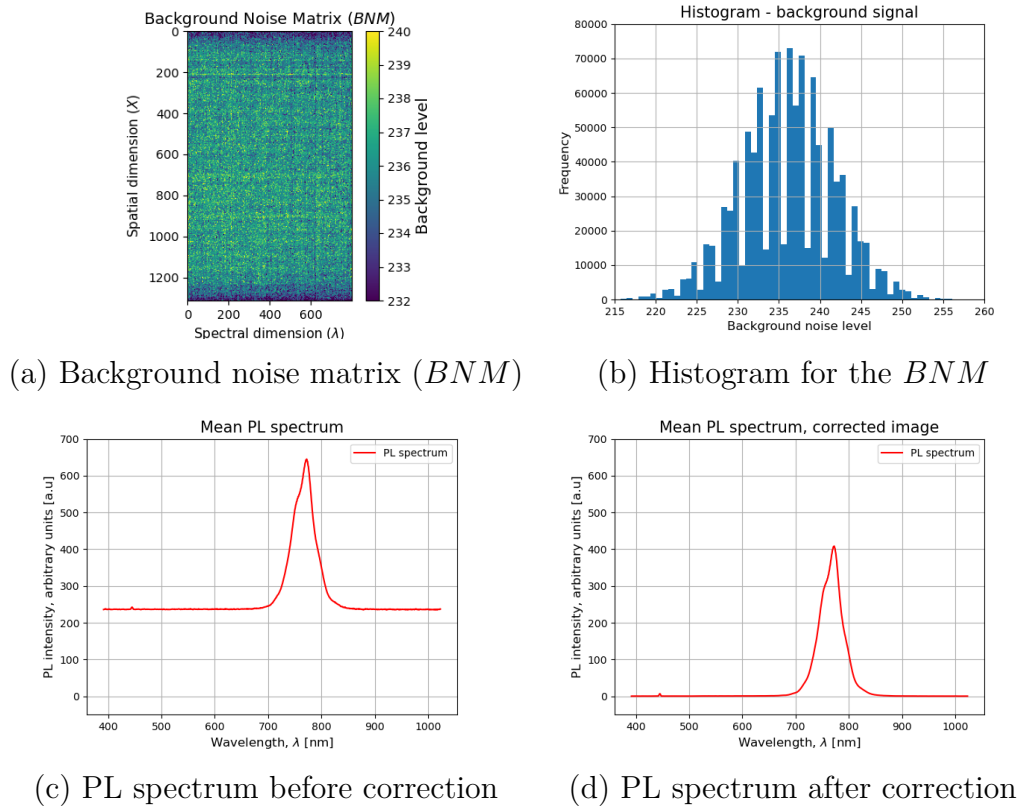


Figure 5.9: *The effect of the background noise correction scheme. (a) Visualization of the background noise matrix ( $BNM$ ) associated with the image of sample 1. The matrix quantifies the background noise in each pixel ( $X, \lambda$ ) on the camera's detector array. (b) The histogram for the  $BNM$ . (c) Mean PL spectrum of the entire sample 1 image before background correction. (d) The mean PL spectrum after background correction.*

Figures 5.9 (c) and (d) compare the sample 1 image's mean PL spectrum before and after performing the background correction. The background correction scheme helps remove the spectrum's vertical offset. Moreover, one can observe that the spectrum also becomes slightly smoother after the correction, since subtracting the  $BNM$  helps remove noise caused by detector variations.

### 5.5.2 Pixel spectra normalization

Figure 5.10 shows the normalization constants (the integrated PL intensities) for each pixel in the HSPL image of sample 1, calculated using Equation (4.4) and the normalization scheme described in Section 4.2.3. This image shows that there are some differences in the total PL intensities among the image's pixels. One can observe some darker and brighter spots, in addition to some scratches and pinholes on the sample. Performing PCA on the image with and without normalization will therefore be tested in Section 5.7.

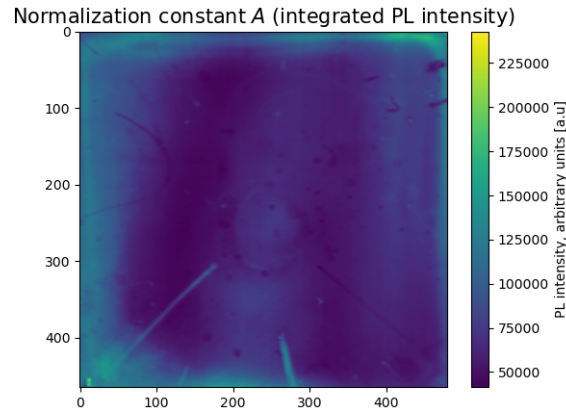


Figure 5.10: *The normalization constants (integrated PL intensities) for sample 1, calculated using Equation (4.4).*

Note that integrating the PL spectra (as in Figure 5.10) produces similar results as summing the counts in all wavelength bands (Figure 5.3). However, by integrating the spectra, the non-uniform widths of the wavelength bands are taken into account. This is why the numerical integration scheme is preferred in this work.

## 5.6 Mean PL Spectrum in the Energy Domain (Sample 1)

The estimation of the perovskite’s bandgap from Section 5.1 (1.60 eV) aligns well with sample 1’s mean PL spectrum shown in Figure 5.11. The PL spectrum has a maximum at 1.61 eV.

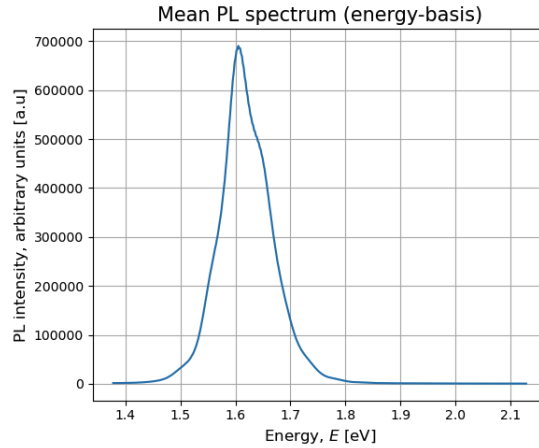


Figure 5.11: *Sample 1’s mean PL spectrum in the energy domain. The spectrum has a peak at 1.61 eV, which is close to the perovskite’s estimated bandgap (1.60 eV). In contrast to the spectra in Figures 5.4 and 5.6, this spectrum is calculated based on a cropped image that only contains the area covered by the sample.*

As can be seen in the mean PL spectrum in Figure 5.11, no PL signal is observed at 2.17 eV, which corresponds to the sample’s bandgap if it only contained bromide ( $y = 1$ ) (see Section 5.1). 1.50 eV, corresponding to  $y = 0$ , however, is part of the sample’s PL peak. It also looks like the PL spectrum has a small shoulder at this energy. Consequently, it may be possible to find regions with only iodide ( $y = 0$ ) in the sample, corresponding to a bandgap at 1.50 eV. Hoke et al. [37] found that the lower-bandgap, iodide-rich domains produce strong PL signals. They explained this by stating that photogenerated charge carriers relax into energy states with the lowest energies, which are found in the lower-bandgap, iodide-rich domains. Much of the radiative recombination consequently occurs in these domains, which gives these regions high luminescence efficiencies. However, the iodide-rich domains they detected had  $y = 0.2$ , and not  $y = 0$  [37]. Therefore, the shoulder at 1.50 eV may originate from another mechanism than BB recombination in a domain with  $y = 0$ . The different components of the PL signal and their origins will be discussed thoroughly in Section 5.9 after MCR has been applied to the image.

## 5.7 PCA (Sample 1)

In this subsection, PCA is performed on the spatial HSPL image of sample 1. PCA is performed both with and without first applying the normalization scheme described in Section 4.2.3.

### 5.7.1 PCA without normalization

#### Scree plot and explained variance

Figure 5.12 shows the scree plot obtained by performing PCA on the HSPL image (without normalization) of sample 1 and subsequently using Equation (4.6) for calculating the explained variance ratios of the principal components. Only the first nine principal components are shown in the plot. PC1 and PC2 explain 98.2% and 1.4% of the total variance in the image, respectively. 0.2% of the variance is explained by PC3, and PC4 explains 0.05% of the variance. The remaining principal components explain less than 0.02% of the variance each. This means that 99.6% of the total variance in the image can be explained by using only two principal components. This represents a significant dimensionality reduction.

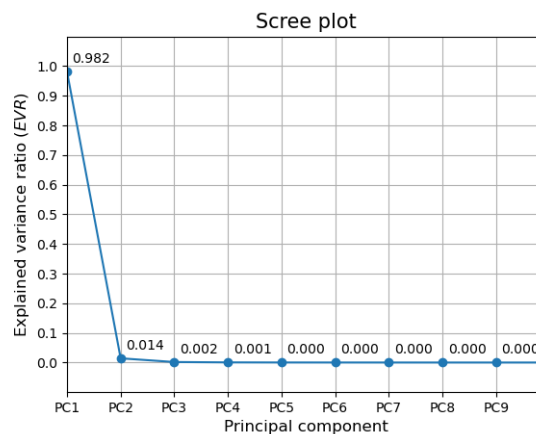


Figure 5.12: *Scree plot for PCA performed on non-normalized sample 1 image. Only the first nine principal components are included in the plot.*

#### PCA scores and loadings

Figure 5.13 shows the score images and corresponding loading plots for the first four principal components (PC1-PC4). The remaining principal components appear noisy, and since they only explain less than 0.02% of the variance each, they are discarded.

## 5.7 PCA (Sample 1)

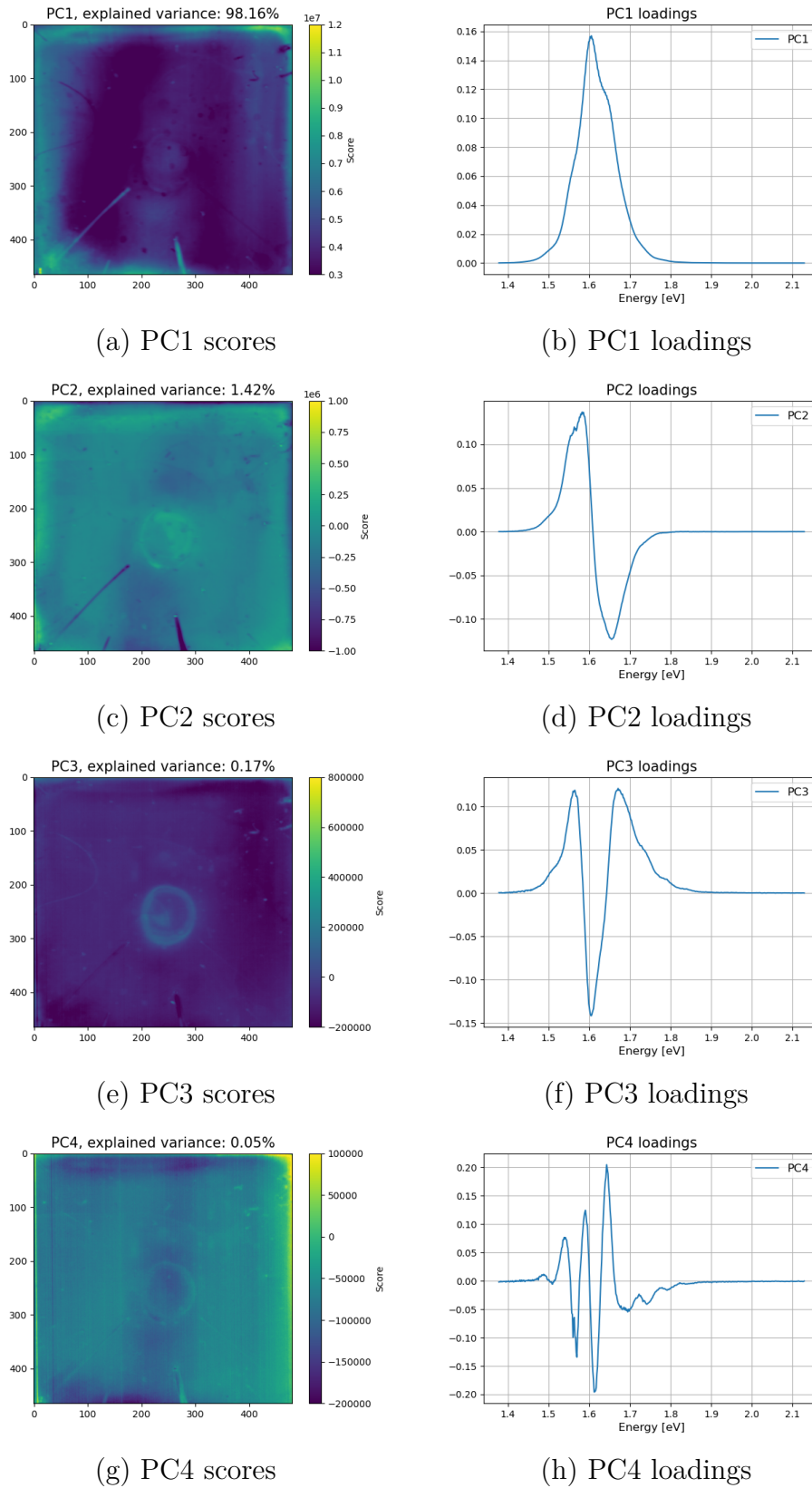


Figure 5.13: Scores and loadings of the first four principal components when PCA is performed on the sample 1 image without normalization.

### Discussion: spectral information extracted by the principal components

Figure 5.14 compares PC1's loading plot (Fig. 5.14 (a)) to the sample's mean PL spectrum (Fig. 5.14 (b)), in addition to comparing the PC1 score image (Fig. 5.14 (c)) to the total PL intensity map (Fig. 5.14 (d)). PC1's loading plot resembles the sample's mean PL spectrum. Moreover, the score image shows the same spatial patterns as the total PL intensity map. Therefore, it seems like PC1 simply captures intensity variations across the sample's surface. PC1 explains 98.2% of the total variance in the image, so these intensity variations are by far the largest contributor to the image's variance when the areas underneath the pixel spectra are not normalized.

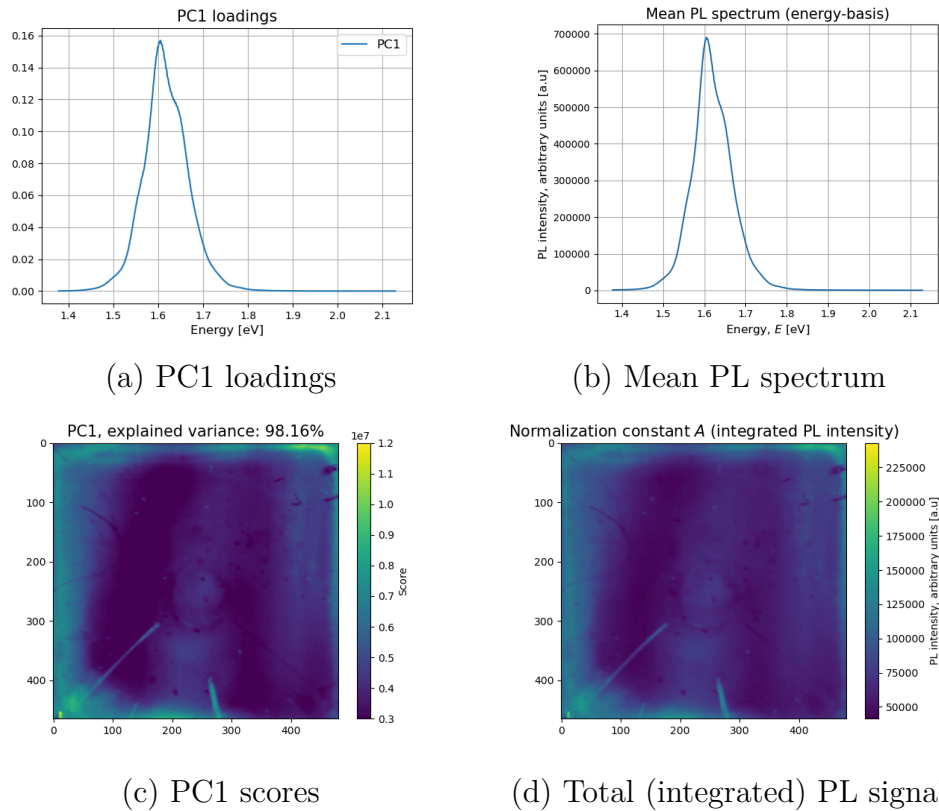


Figure 5.14: *Interpreting the first principal component (PC1).* Figure (a) shows the PC1 loadings, and (b) shows sample 1's mean PL spectrum. Figure (c) shows the PC1 score image and Figure (d) shows the total (integrated) PL signal (see Figure 5.10). The similarities between (a) and (b), and (c) and (d), suggest that PC1 extracts information about intensity variations across the sample's surface.

PC2's loading plot (Fig. 5.13 (d)) has a maximum at 1.58 eV and a minimum



at 1.65 eV. PC2, therefore, distinguishes between regions that emit PL signals at different energies. Pixels with positive PC2 scores emit strongly at around 1.58 eV, and pixels with negative PC2 scores emit PL signals at around 1.65 eV. Since the perovskite's bandgap depends on the bromide-to-iodide ratio (see Section 5.1), these spectral shifts may be caused by phase segregation-related bandgap variations. If the spectral shifts are related to phase segregation, the maximum at 1.58 eV may represent regions with higher iodide concentrations, and the minimum at 1.65 eV may represent regions with higher bromide concentrations. Consequently, pixels with positive or negative PC2 scores may correspond to regions with a lower or higher bromide-to-iodide ratio, respectively. Note also that the PC2 loading plot has a bump at around 1.5 eV, which may be caused by BB emissions from iodide-rich domains with  $y$  close to zero or by other recombination routes (see Section 5.6).

PC3's loading plot (Fig. 5.13 (f)) has two maxima, one at 1.56 eV and another at 1.68 eV. A positive PC3 score indicates that a pixel emits PL signals in the spectral regions around these energies. These two maxima are close to PC2's extremes (at 1.58 eV and 1.65 eV), which means that they most likely carry similar information as PC2. The peak at 1.56 eV may represent regions with higher iodide concentrations, and the peak at 1.67 eV may represent regions with higher bromide concentrations. The same bump at 1.50 eV, as was seen in PC2's loading plot, can also be observed in PC3's loadings. The loading plot also has a minimum at 1.61 eV, which coincides with the sample's PL maximum (see Figure 5.11). This minimum is also close to 1.60 eV, which is the bandgap estimated based on the linear regression model from Li et al. [32] (Equations (2.1) and (2.2)). The minimum at 1.61 eV in PC3's loading plot may therefore represent a mixed, unsegregated phase, where the local halide stoichiometry is equal to the sample's overall halide stoichiometry. A possible interpretation of PC3 is, therefore, that it distinguishes between a mixed phase (negative PC3 scores) and regions where the stoichiometry is different (positive PC3 scores), with either a higher or a lower bromide-to-iodide ratio than the mixed phase.

The loading plot for PC4 (Fig. 5.13 (h)) oscillates. Moreover, the score plot (Fig. 5.13 (g)) has two prominent vertical lines on its left-hand side, one bright and one dark line. Hence, it looks like PC4 captures some detector variations that are not accounted for by the background correction scheme. Similar stripes in the score images and oscillations in the loading plots are also observed for the subsequent principal components. PC4 and the remaining principal components are therefore less relevant than PC2 and PC3 regarding

phase segregation.

### 5.7.2 PCA with normalization

#### Scree plot and explained variance

Figure 5.15 shows the scree plot obtained when performing PCA on the normalized hyperspectral image of sample 1. The plot shows that PC1 now explains 67.8% of the total variance in the image. PC2, PC3, and PC4 capture 15.2%, 4.17%, and 0.831% of the variance, respectively. The remaining principal components explain less than 0.5% each.

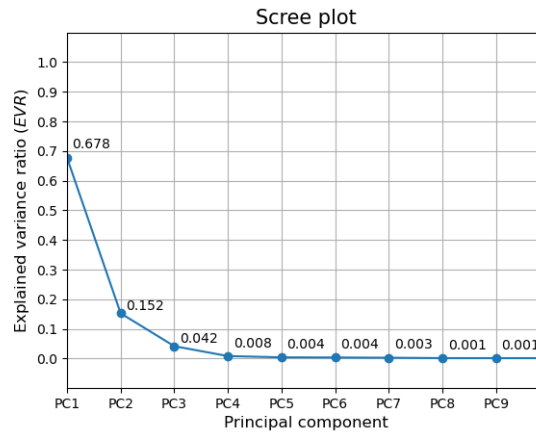


Figure 5.15: *Scree plot for PCA performed on the normalized sample 1 image.*

#### PCA scores and loadings

Figure 5.16 shows the score images and corresponding loading plots for the first four principal components (PC1-PC4) when the pixel spectra are normalized.

#### Discussion: spectral information extracted by the principal components

PC1's loading plot (Fig. 5.16 (b)) has a maximum at 1.58 eV and a minimum at 1.65 eV. These extremes coincide with the extremes in the PC2 loading plot produced by performing PCA on the non-normalized image (see Section 5.7.1). Moreover, since these two loading plots are very similar, it seems like PC1 in the normalized image captures the same information as PC2 in the non-normalized image. Positive PC1 scores in Figure 5.16 (a) may therefore represent iodide-rich regions, and negative PC1 scores correspond to bromide-rich regions.

The PC2 loading plot in Figure 5.16 (d) also looks very similar to the PC3 loading plot in Figure 5.13 (f). The two loading plots have two maxima, located at 1.56 eV and 1.68 eV, in addition to a minimum at around 1.60 eV. The only noticeable difference between these loading plots is that the peak at 1.56 eV is less intense in Figure 5.16 (d) than in Figure 5.13 (f), relative to the peak at 1.68 eV. However, it seems like PC2 in the normalized image captures much of the same information as PC3 in the non-normalized image. PC2 hence seems to distinguish between regions with the same halide stoichiometry as the sample's overall halide stoichiometry and regions with bromide-to-iodide ratios that differ from this.

The PC3 and PC4 loading plots in Figures 5.16 (f) and (h) oscillate, and are hence difficult to relate to phase segregation. Moreover, their corresponding score images in Figures 5.16 (e) and (g) have similar vertical stripes as the PC4 score image in Figure 5.13 (g). Therefore, it is assumed that these components capture detector variations that are not corrected by the background correction scheme.

## 5.7 PCA (Sample 1)

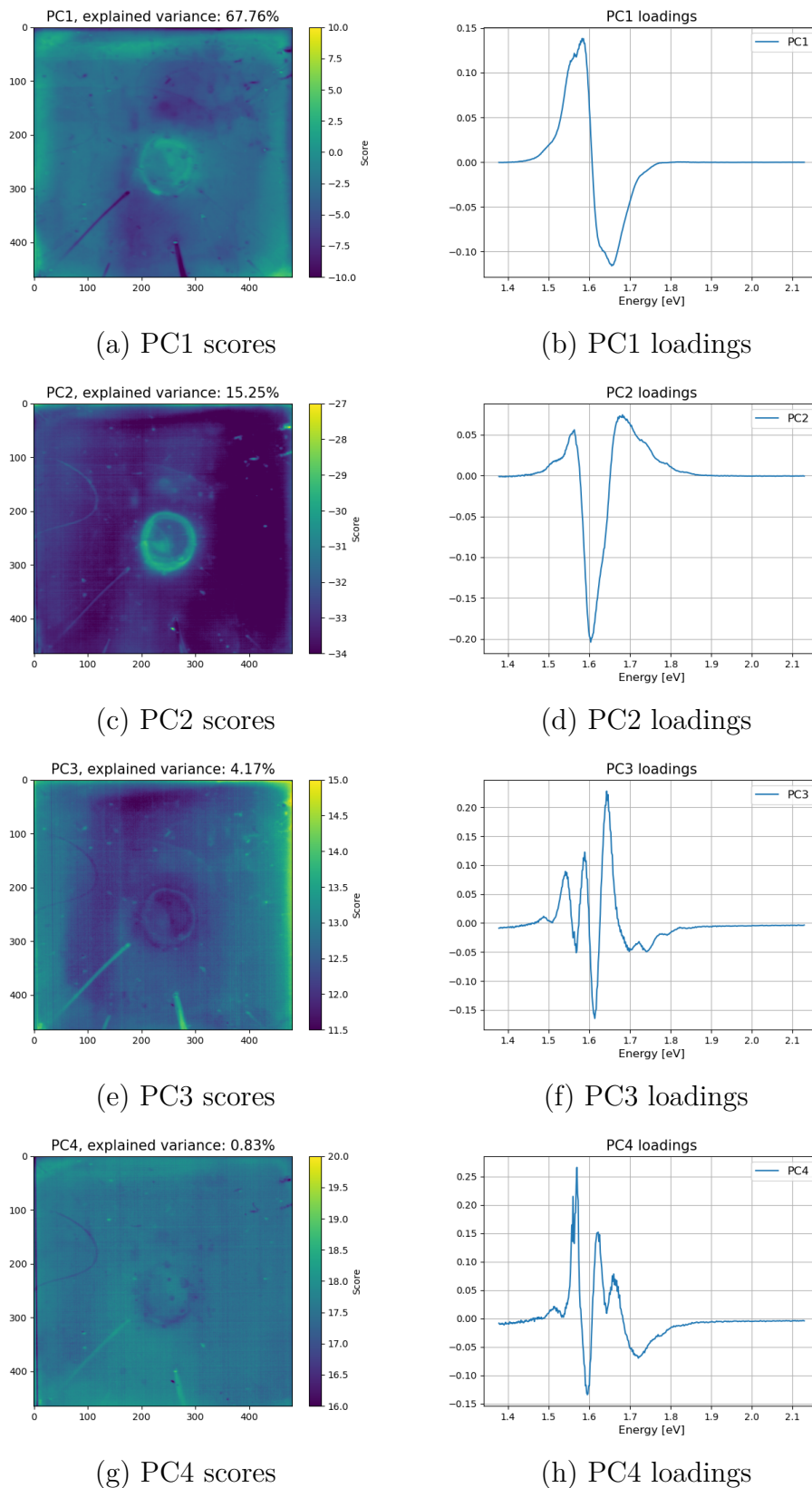


Figure 5.16: Scores and loadings of the first four principal components when PCA is performed on the normalized sample 1 image.

### 5.7.3 Discussion: comparing PCA with and without normalization

Based on the observations made in Sections 5.7.1 and 5.7.2, the major difference between applying the normalization scheme or not before performing PCA is whether the model captures PL intensity variations or not. When the image is not normalized before performing PCA, the first principal component captures the PL intensity variations across the sample's surface. PC2 and PC3 capture information that may be relevant for mapping phase segregation. On the contrary, when the image is normalized, PC1 and PC2 capture similar information as PC2 and PC3 did for the non-normalized image. Consequently, the results are similar whether the normalization scheme is applied or not, but the normalization scheme can be used to exclude the PL intensity variations from the analysis.

### 5.7.4 Discussion: spatial information extracted by the principal components

#### Regions with high or low PC2 scores (non-normalized image)

For examining the spatial information captured by PC2 when performing PCA on the non-normalized image, four image regions ( $10 \times 10$  pixels) are selected by segmenting out pixels with high or low PC2 scores. Binary images created by thresholding at two different PC2 score levels ( $> 200000$  and  $< -400000$ ) are shown in Figures 5.17 (a) and (b). Regions A and D with large, positive PC2 scores, and regions B and C with large, negative PC2 scores are selected based on these thresholded images. These regions are marked in the PC2 score image and the total PL intensity map in Figures 5.17 (c) and (d), respectively.

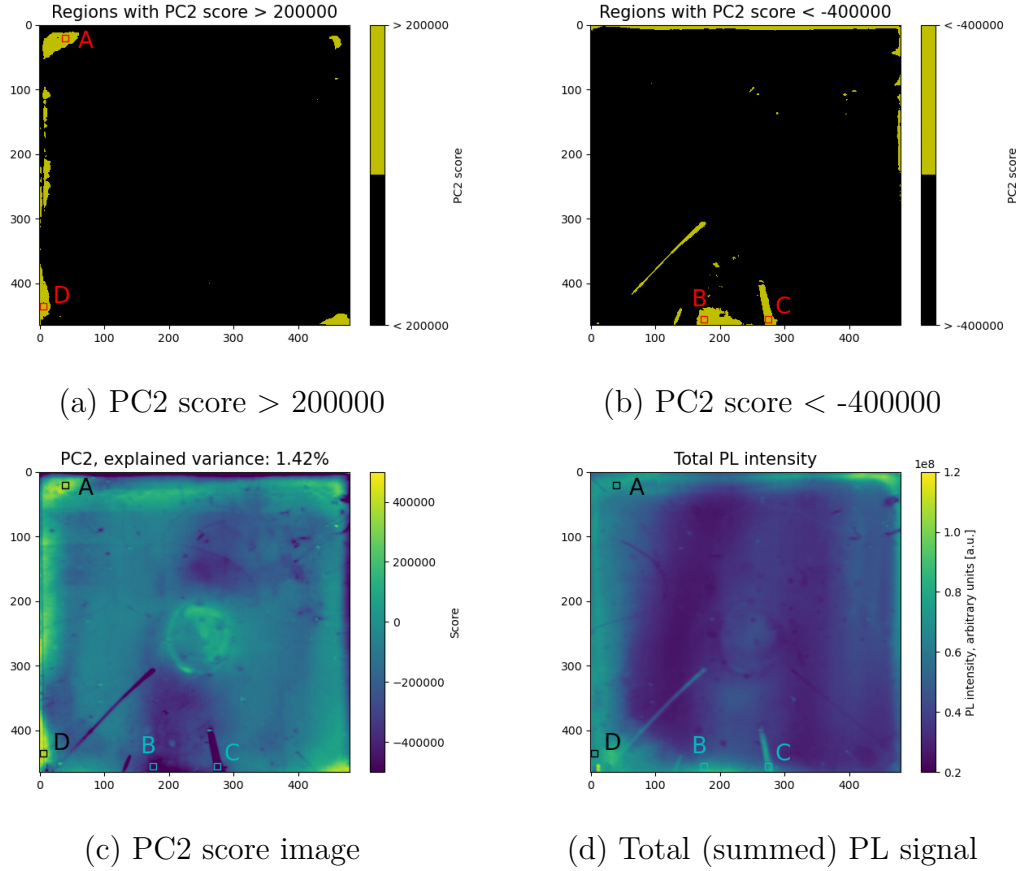


Figure 5.17: Selection of regions with high or low PC2 scores (non-normalized image). Figures (a) and (b) show binary images where pixels with high or low PC2 scores have been segmented out. Regions A and D have large, positive PC2 scores, while regions B and C have large, negative PC2 scores. Figures (c) and (d) show these regions marked in the PC2 score image and the total PL intensity (sum of bands) image, respectively.

The four regions' mean PL spectra are calculated and compared in Figure 5.18. As expected, regions A and D have very similar spectra (see Fig. 5.18 (a)), as these spectra originate from pixels with similar PC2 scores. The same applies to regions B and C (see Fig. 5.18 (b)). However, when regions with different PC2 scores are compared, some differences can be noticed. Region D's mean PL spectrum is compared to the mean spectra of regions B and C in Figures 5.18 (c) and (d). A small shift along the energy axis in the PL spectra can be observed in Figure 5.18 (c), but the shift is slightly larger in Figure 5.18 (d). The location of the PL intensity maximum changes only a little between regions C and D, but there are some differences in the intensities especially at around 1.55 eV and 1.65 eV. Spectra from regions B and C are

blueshifted relative to the spectra from regions A and D, which may be caused by differences in the bromide-to-iodide ratios.

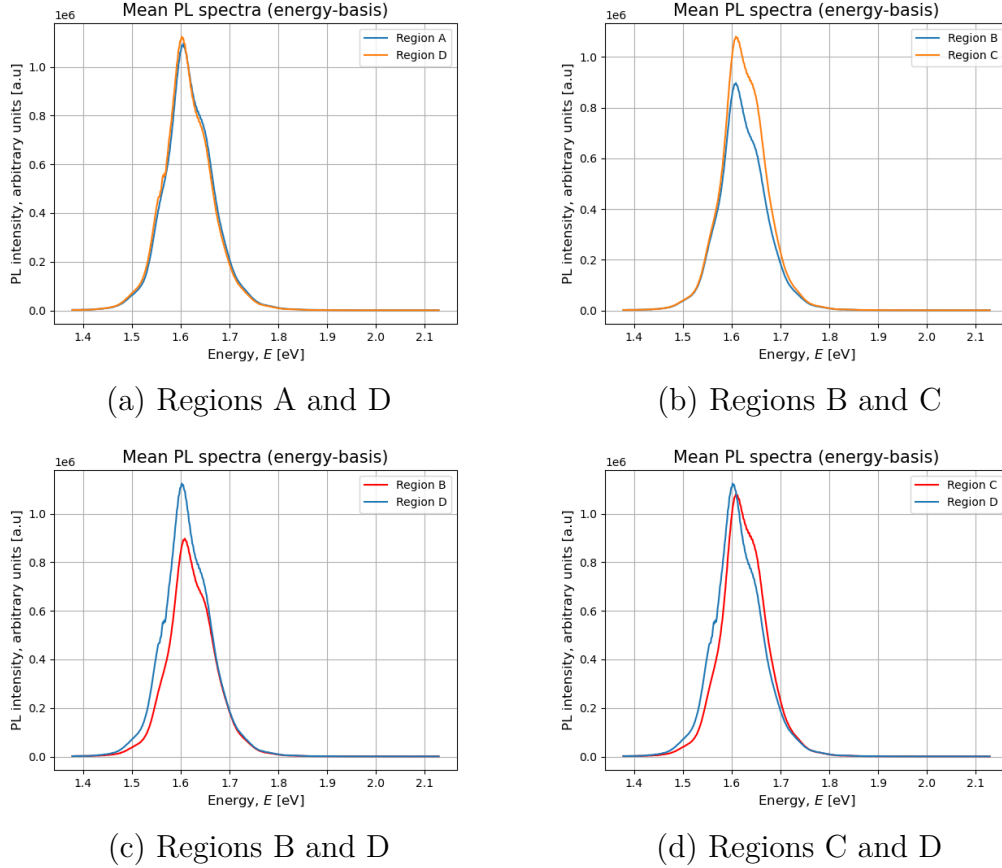


Figure 5.18: *Pair-wise comparisons of the mean PL spectra of the selected regions from Figure 5.17.*

One can observe that all spectra in Figure 5.18 have a PL intensity shoulder at an energy higher than that of the PL peak. This suggests that all four regions may contain higher-bandgap, bromide-rich domains, in addition to a mixed phase corresponding to a bandgap at approximately 1.60 eV. The four spectra also comprise the previously mentioned shoulder at around 1.50 eV. Moreover, region D has an intensity shoulder at around 1.57 eV, which could potentially be attributed to domains of higher iodide concentrations. This shoulder is not as prominent in the other spectra, but it looks like the other spectra may have a small shoulder at around 1.55-1.57 eV. Therefore, based on the regions' mean spectra, it is expected that the different phases (iodide-rich, mixed, and bromide-rich) are present to a greater or lesser extent in all four regions.

The shoulder at around 1.55-1.57 eV seems a bit noisy in region D. As can be

seen in Figure F.1 in Appendix F, this is explained by noise in some of region D's pixel spectra.

### 5.7.5 Summary of the PCA analysis

As can be seen in Figures 5.11, 5.18 and F.1, PL spectra for the entire sample, smaller regions, and even single pixels have intensity shoulders and bumps on both sides of the PL intensity maximum. This indicates that the total PL signals might be a superposition of a set of fundamental signals for instance originating from domains of different halide stoichiometries. Barker et al. [67] state that iodide-rich regions typically form close to the illuminated surface of the perovskite material. The composite PL signals in single pixels may therefore be a superposition PL signals originating from different depths of the perovskite film. MCR aims to decompose pixel spectra into their constituents and may bring additional insight.



## 5.8 Exploring and Assessing Different MCR Models (Sample 1)

This subsection deals with the development of an MCR model for the HSPL image of sample 1. The next subsection (Section 5.9) will focus on interpreting the output from this model.

The non-normalized image of sample 1 is used as input to the MCR algorithm. Therefore, instead of using PC1 to generate the spectral guesses as was done for the normalized EL image in Section 4.3.2, the PC2 peaks (Figure 5.13 (d)) are used here. Apart from this, the procedure remains unchanged.

### 5.8.1 Choosing the number of MCR components

#### MCR without regularization (MCR-ALS)

The "lack of fit" ( $LoF$ ) values can help determine the appropriate number of MCR components, as discussed in Section 4.3.2. The  $LoF$  is first calculated for a set of non-regularized MCR models (i.e. by using the default OLS regressor). In Figure 5.19 (a), the  $LoF$  is plotted as a function of the number of MCR components used in these models. Figure 5.19 (b) zooms in and shows only the models with less than 10 components. The figure shows that the minimum  $LoF$  value, which is 1.36%, is obtained when using four components. This means that according to this metric, the non-regularized model using four MCR components reconstructs the pixel spectra more accurately than the rest of the investigated models. Moreover, the models having three, five, six, or seven components all have  $LoF$  values below 2%.

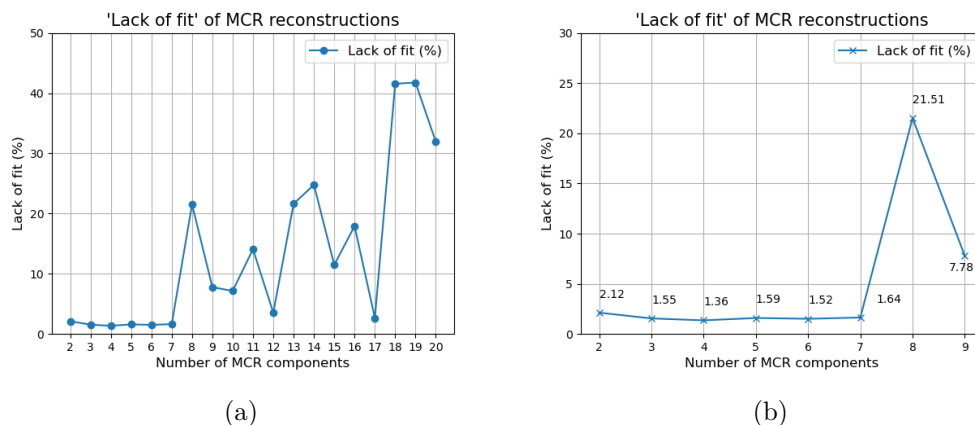


Figure 5.19: "Lack of fit" (LoF) of selected non-regularized MCR models. The LoF values are plotted as a function of the number of MCR components. Models ranging from two to 20 components were investigated in (a). In (b), only the models having less than ten components are included. The LoF values are written next to each data point for better readability.

Pure spectra of the non-regularized models with between two and seven MCR components are shown in Figure 5.20. The figure shows that the second MCR component in the two-component model is resolved into two peaks when a third component is added. Therefore, it is likely that a model with three components describes the system better than a model with only two components. The figure also shows that in the four-component model, three of the components are similar to the components of the three-component model, but an additional component with several peaks is added. This fourth component may be a consequence of the model adapting to noise, so a thorough analysis has to be conducted before it is concluded to use this number of components. When more components are added, some of the pure spectra contain only zeros or very small values compared to the rest of the spectra. The next paragraphs discuss the problems that may arise when using too many components and when performing non-regularized regression.

### Discussion: risk of overfitting by using too many components

In addition to the reconstruction error, the risk of overfitting also has to be taken into consideration when determining the number of MCR components. If many components are used, the model will comprise a large number of trainable parameters, which may result in a model that is more complex than desirable [53, p. 75]. The model may reproduce the original spectra well, but there is a risk that this is also a consequence of the model adapting to the

## 5.8 Exploring and Assessing Different MCR Models (Sample 1)

noise present in the data. Therefore, a model with fewer components may be more successful in extracting just the relevant information from the dataset. Moreover, a model comprising fewer components is easier to interpret, since it explains the system in simpler terms. These arguments support exploring models with fewer components first, and then only adding more components if that seems to enhance the analysis. One has to critically evaluate whether the models with, for instance, six or seven components explain more of the underlying physics than the simpler models with fewer components.

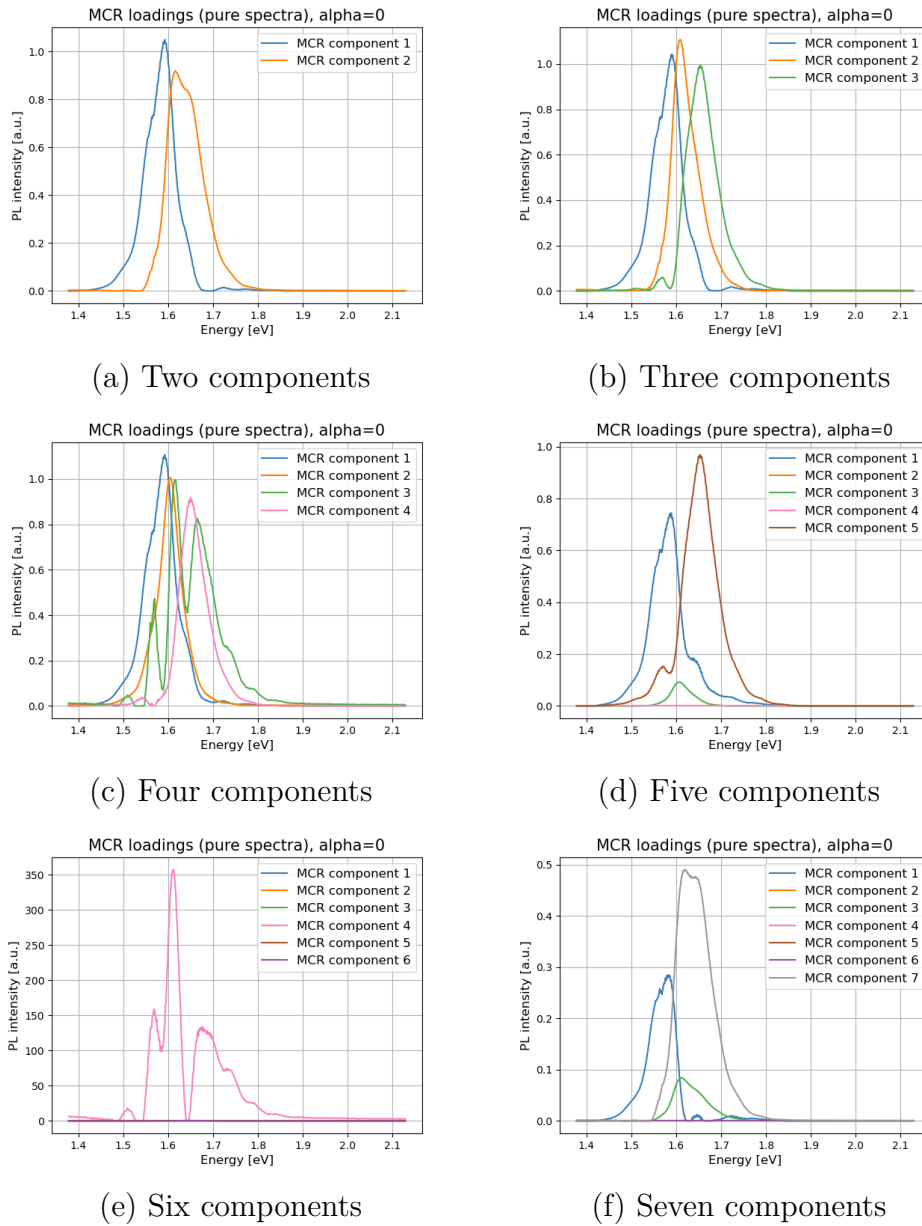


Figure 5.20: *Pure spectra of the non-regularized MCR models with between two and seven components.*

**Discussion: benefits of adding a regularization term**

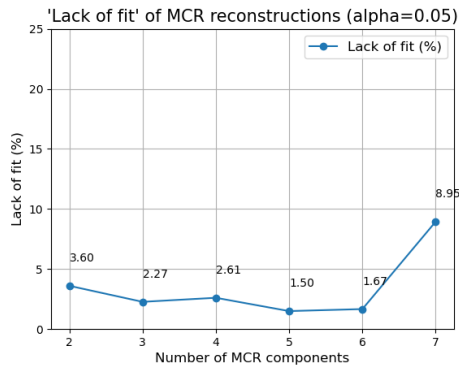
Figure G.1 in Appendix G motivates why it may be beneficial to penalize large model weights by introducing a regularization term. The figure shows selected score plots (concentrations) and the pure spectra of the non-regularized MCR model with six components. The model has a low reconstruction error (a *LoF* value of 1.52% in Figure 5.19 (b)), and only one model achieves a lower error. However, as can be seen in Figures G.1 (a)-(d), the model's concentration values vary by many orders of magnitude between the components. This also applies to the values of the pure spectra (even though this is difficult to see in the figure since some spectra have very small values). Some of the spectra also look very noisy. Hence, the model seems to overfit by adapting to noise in the data. A situation like this can be prevented by building a simpler model, either by using fewer components or by penalizing extreme model weight values (concentrations and pure spectra). A regularization term will penalize the model for having large weights, and this may prevent overfitting (see Section 2.6.2).

In addition to motivating the need for a regularization term, the aforementioned model also highlights the importance of considering more than the *LoF* value when choosing an MCR model. When adding more components, the key question is whether the more complex model explains more of the underlying physics or if it achieves a lower reconstruction error by adapting to noise in the image.

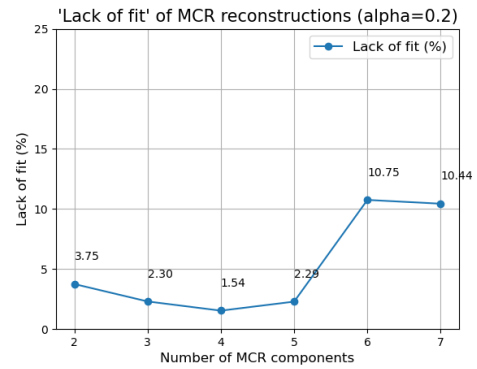
**MCR with regularization (Ridge Regression)**

Regularization is tested on MCR models with between two and seven MCR components since these models yielded the lowest reconstruction errors in Figure 5.19. Figure 5.21 shows the *LoF* as a function of the number of MCR components for selected values of the regularization parameter  $\alpha$ .

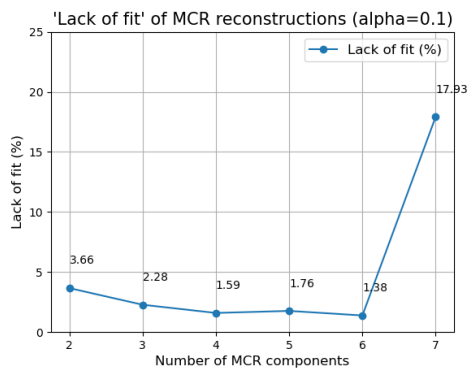
## 5.8 Exploring and Assessing Different MCR Models (Sample 1)



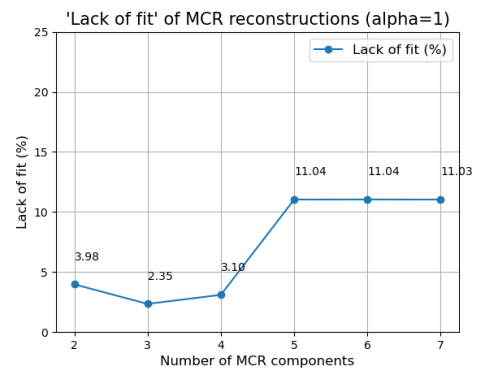
(a)  $\alpha = 0.05$



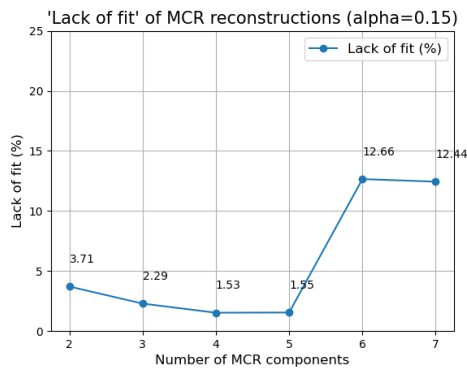
(d)  $\alpha = 0.2$



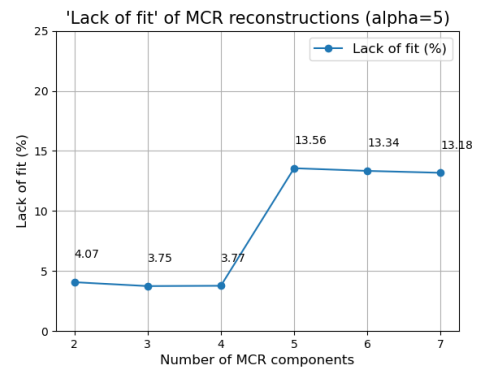
(b)  $\alpha = 0.10$



(e)  $\alpha = 1$



(c)  $\alpha = 0.15$



(f)  $\alpha = 5$

Figure 5.21: "Lack of fit" (LoF) as a function of the number of MCR components for different values of the regularization parameter  $\alpha$ .

Several observations can be made in Figure 5.21. First, the reconstruction error is lower when using three components than when using two components for all values of  $\alpha$ . Since the reconstruction error is lower also for the higher values of  $\alpha$ , the increase in performance is most likely not a consequence of overfitting. Therefore, Figure 5.21 suggests that adding the third component helps the

model describe the system better. This aligns well with the observation made in Figure 5.20, namely that the second MCR component is resolved into two components when going from two to three components.

Second, the number of MCR components that yield the lowest reconstruction error varies with the value of the regularization parameter  $\alpha$ . However, the models with three or four components often minimize the error.

Third, as the value of  $\alpha$  increases, the models with many components tend to perform poorer than the models with fewer components. One possibility is that the models with many components perform well for small values of  $\alpha$  due to overfitting. When  $\alpha$  is small, their large weights are not penalized. However, when the regularization strength increases, these complex models are penalized for having large weights. The result is that they no longer succeed in reconstructing the image's original pixel spectra.

A plot of  $LoF$  as a function of  $\alpha$  for MCR models with different numbers of components is given in Figure G.2 in Appendix G. This figure demonstrates something that can also be observed in Figure 5.21, namely that the models with fewer components are less influenced by the value of the regularization parameter. These models are, therefore, less likely to overfit by having extreme model weights. The more complex models are less stable and more influenced by how strongly they are penalized by the regularization term.

Based on Figures 5.21 and G.2, it is decided to investigate the MCR models with three, four, and five components further. The model with two components is excluded since one of its components is resolved into two components when using a model with three MCR components (see Figure 5.20). Moreover, the spectra of selected sample regions investigated in Section 5.7.4 suggest that more than two components are present. Recall that the PL spectra, both on the sample- and on the pixel level, had shoulders and bumps that may be attributed to PL signals emitted from phases with different halide stoichiometries. Furthermore, the models with three, four, or five MCR components are preferred over the more complex models, as the models with fewer components seem more robust to changes in the regularization strength.

The value of the regularization parameter  $\alpha$  is set to 0.15, which, based on Figure G.2, seems like a reasonable tradeoff between preventing overfitting by penalizing extreme weights and giving the regression models the freedom to learn from the dataset.

### 5.8.2 Comparing the candidate MCR models

Figure 5.22 shows the pure spectra associated with both non-regularized and regularized MCR models with three, four, and five components. As can be seen in Figures 5.22 (a) and (b), the shapes of the pure spectra for the model with three components are very similar whether the regularization term is added or not. However, for the model with four components (Figures 5.22 (c) and (d)), the regularization strength influences the shape of the third component. With  $\alpha = 0$ , this component has several maxima and covers a large spectral range. The number of maxima is reduced and the spectral range of the component is made narrower when  $\alpha = 0.15$ . This way, the information described by each component becomes more distinct with regularization. The model with four components hence seems more sensible with regularization than without.

The model with five components has some issues both with and without regularization. Without regularization (Figure 5.22 (e)), component 2's spectrum is just a scaled version of component 1's spectrum (this is not possible to see in the figure because the second component's values are approx.  $10^{-14}$  of the first component's values). Moreover, the component 4 spectrum consists of only zeros. Since the remaining three components have many of the same spectral features as the components in the MCR model with three components, it is concluded that the non-regularized MCR model with five components adds nothing to the analysis. With  $\alpha = 0.15$ , several of the spectra appear noisy. Components 2 and 4 have spectra comprising several maxima. Moreover, sharp peaks are observed at around 1.56-1.57 eV in the spectra of components 1 and 3. These peaks may be caused by the same noise as was seen in the pixel spectra in Figure F.1. Vertical stripes are also observed in the model's score images (see Figure G.5 in Appendix G), which supports that the model is affected by noise that is not accounted for by the background correction scheme.

## 5.8 Exploring and Assessing Different MCR Models (Sample 1)

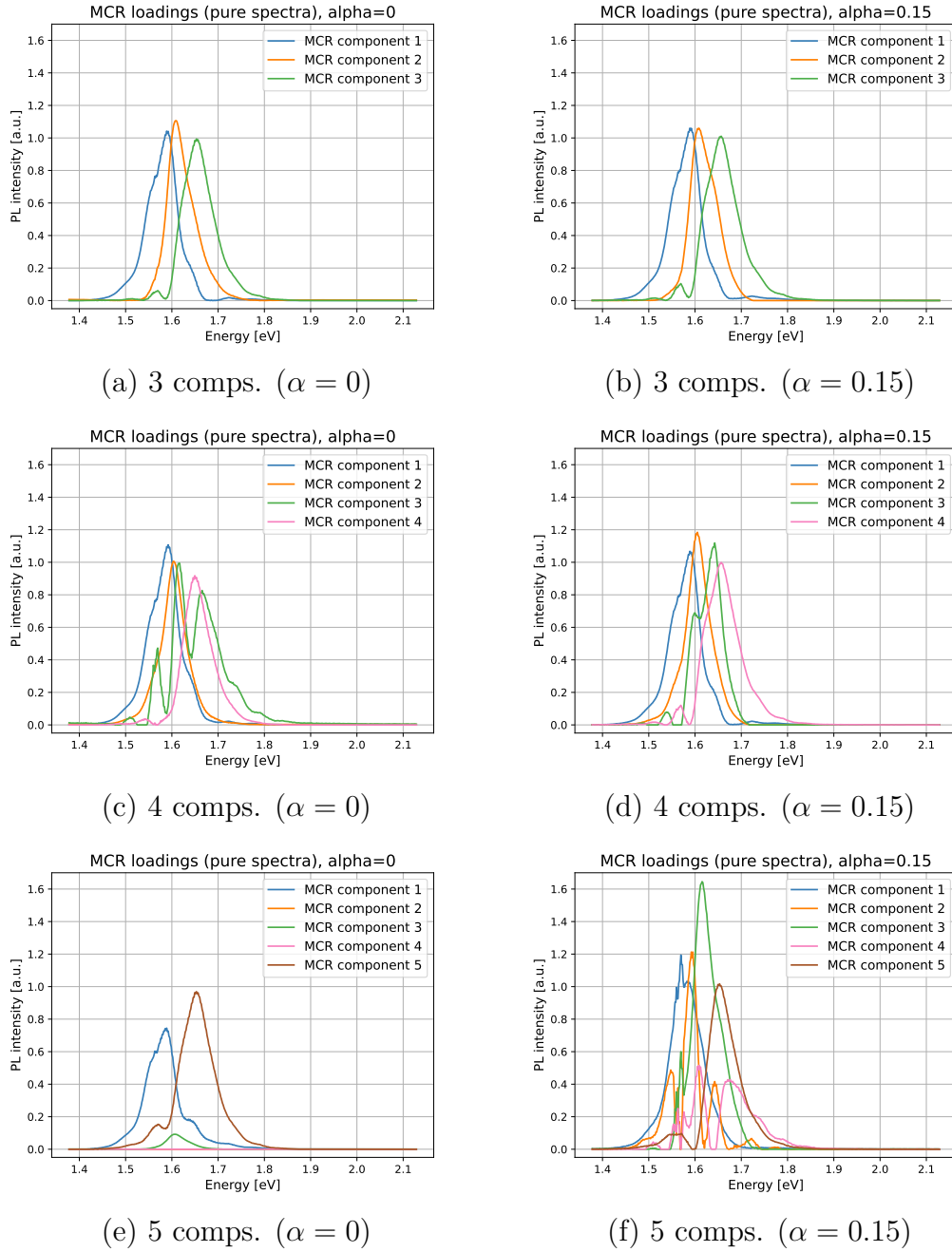


Figure 5.22: *Pure spectra of MCR models with three, four, and five components, both with and without regularization (corresponding to  $\alpha = 0.15$  and  $\alpha = 0$ , respectively).*

Based on the discussion above, it is decided to examine the three-component model with  $\alpha = 0$  (Figure 5.22 (a)) and the four-component model with  $\alpha = 0.15$  (Figure 5.22 (d)) further. Score images and pure spectra for these two models can be found in Figures G.3 and G.4 in Appendix G. Regularization



is omitted for the three-component model since the spectra are very similar regardless of whether regularization is applied or not. Because the model has only three components, it is less likely than the more complex models to suffer from overfitting. Therefore, minimizing the model's reconstruction error by setting  $\alpha = 0$  is prioritized. The four-component model, on the other hand, seems to benefit from regularization, as discussed earlier. Also, the locations of the pure spectra's PL peaks are identical for different values of  $\alpha$  larger than zero, as can be seen in Figure G.6 in Appendix G. Therefore, the pure spectra for different regularization strengths carry similar physical information. Using  $\alpha = 0.15$ , which represents a local *LoF* minimum in Figure G.2 (c), seems sensible.

It can not be ruled out that more than four components are present in the system. The pure spectra in Figure 5.22 often have more than one maximum each. Therefore, these spectra should ideally have been decomposed into more components. However, it has proven difficult to build MCR models with more components without generating noisy pure spectra. It may be that the different peaks are difficult to resolve because they are highly correlated, in the sense that they can be found in the same pixels/sample regions. This will be discussed more in Section 5.9.1.

### 5.8.3 Pixel-wise "lack of fit" (*PLoF*) of candidate models

The non-regularized three-component model and the regularized four-component model have similar *LoF* values of 1.55% and 1.53%, respectively (see Figure G.2). However, Figure 5.23 shows that the models yield reconstruction errors that are somewhat differently distributed among the pixels. The figure shows maps and histograms of the pixel-wise "lack of fit" (*PLoF*) values calculated by using Equation (4.8). As can be seen by comparing Figure 5.23 (a) and (b), the three-component model has a higher reconstruction error in the circular region in the center of the sample and in the stripe at the bottom of the sample (where region C is). The distribution of the *PLoF* values is also narrower in Figure 5.23 (d) than in Figure 5.23 (c). This means that the reconstruction errors of the four-component model are more homogeneous across the sample's surface.

Figures 5.23 (a) and (b) also show that both regions have vertical stripes with higher reconstruction errors on the left-hand side of the sample. These stripes are most likely caused by noise from the detector array that is not accounted for by the background correction scheme. Therefore, it is not problematic that the reconstruction error is higher in these pixels.

## 5.8 Exploring and Assessing Different MCR Models (Sample 1)

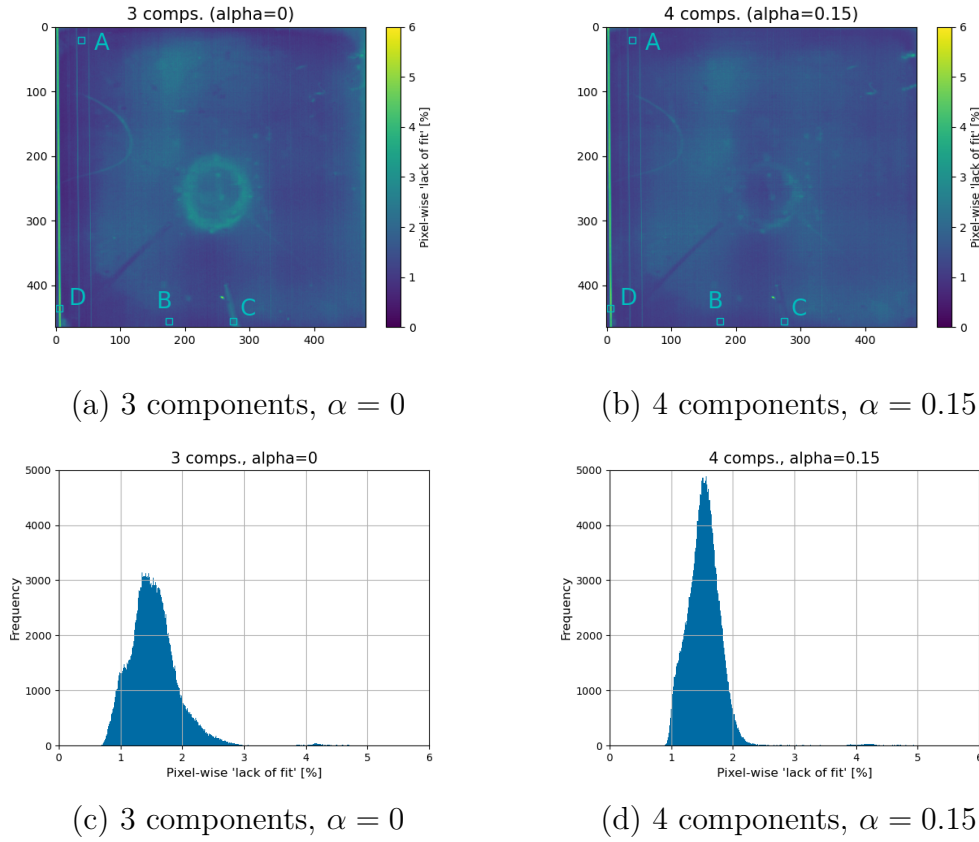


Figure 5.23: *Pixel-wise "lack of fit" (PLoF) maps and histograms for the non-regularized three-component model and the regularized four-component model. Regions A, B, C, and D from Section 5.7.4 are marked in the PLoF maps in (a) and (b).*

### 5.8.4 Comparing the candidate models by decomposing the PL spectra for selected regions

#### Regions A, B, C, and D

The non-regularized MCR model with three components and the regularized model with four components are used to decompose and reconstruct the mean PL spectra of regions A, B, C, and D that were examined in Section 5.7.4. Figure G.7 in Appendix G shows plots of the resulting components and the reconstructed PL spectra of the four regions. The concentrations of each component are calculated as averages over all pixels in the respective regions. The reconstructions are computed as a superposition of the models' pure spectra, using the mean concentrations as weights.

Figure G.7 shows that the two models are able to reconstruct the PL spectra

equally well in regions A, B, and D. In region C, however, the model with three components has a slightly higher reconstruction error than the model with four components, as can also be seen in Figures 5.23 (a) and (b). Figure 5.24 compares the two models in region C and shows that the model with three components does not replicate the PL spectrum’s intensity shoulder at around 1.65 eV as accurately as the model with four components. However, the differences are marginal.

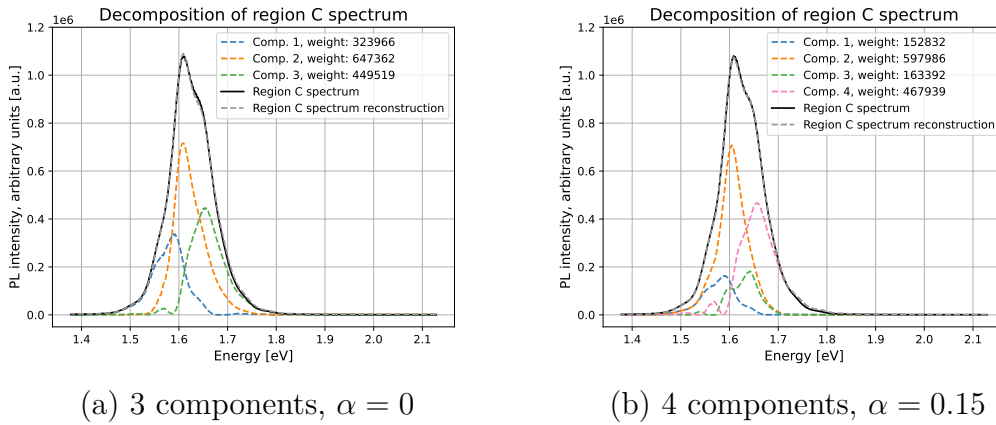


Figure 5.24: Comparing the decompositions and reconstructions of region C’s mean PL spectra obtained using the MCR models with (a) three and (b) four components. The pure spectra are weighted by their respective concentrations, and the reconstructions are the sum of these weighted pure components.

Since the deviation between the reconstructed and original spectrum in Figure 5.24 is small for both models, the additional MCR component does not necessarily enhance the analysis. Component 3’s score image (Figure G.4 in Appendix G) shows that region C is one of the regions where this component’s concentration is the largest. However, as one can see in Figures G.7 (b), (d), (f), and (h) in Appendix G, component 3 is not dominant in any of the regions A, B, C, or D. Consequently, the reconstruction changes minimally whether this component is added or not. Moreover, the component has four different PL peaks, and some of them are close to peaks in the other components. These factors suggest that the purpose of component 3 is to minimize the reconstruction error by enabling minor adjustments to the reconstructed spectra. If the additional component explains more of the physics of the system remains unclear.

### Circular region at the sample's center

As was shown in Figures 5.23 (a) and (b), the MCR model with three components has a larger reconstruction error than the four-component model in the circular region at the sample's center. The circle is most likely where the antisolvent was dropped onto the substrate during spin-coating deposition of the perovskite layer (see Section 5.3.2). Figure 5.25 shows a binary image in which the pixels with  $PLoF$  values between 2.5% and 3.0% for the three-component model are distinguished from those with other  $PLoF$  values. Based on the figure, it is evident that most pixels in the circular region have a  $PLoF$  value between 2.5% and 3.0% for the three-component model.

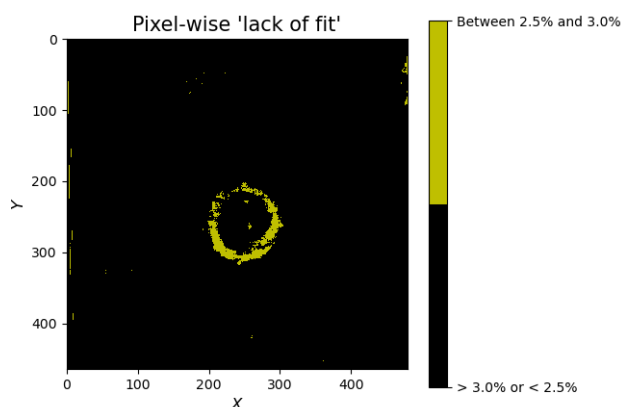


Figure 5.25: *Thresholded pixel-wise "lack of fit" ( $PLoF$ ) map showing the pixels where the reconstruction error of the three-component model is between 2.5% and 3.0%. The circular region at the center of the sample is most likely where the antisolvent was dropped onto the substrate during spin-coating deposition of the perovskite layer (see Section 5.3.2).*

Figure 5.26 compares how the MCR models with three and four components decompose and reconstruct three selected pixel spectra in the circular region. The figure shows that the  $PLoF$  is lower for the four-component model than for the three-component model in all three pixels. However, the three-component model's reconstruction error is, in general, small. Only minor deviations between the reconstructed and original pixel spectra can be observed at around 1.65 eV and between 1.7 eV and 1.8 eV in Figures 5.26 (a)-(c). Observe also that the concentrations of the four-component model's component 3 in Figures 5.26 (d)-(f) are small in all pixels.

## 5.8 Exploring and Assessing Different MCR Models (Sample 1)

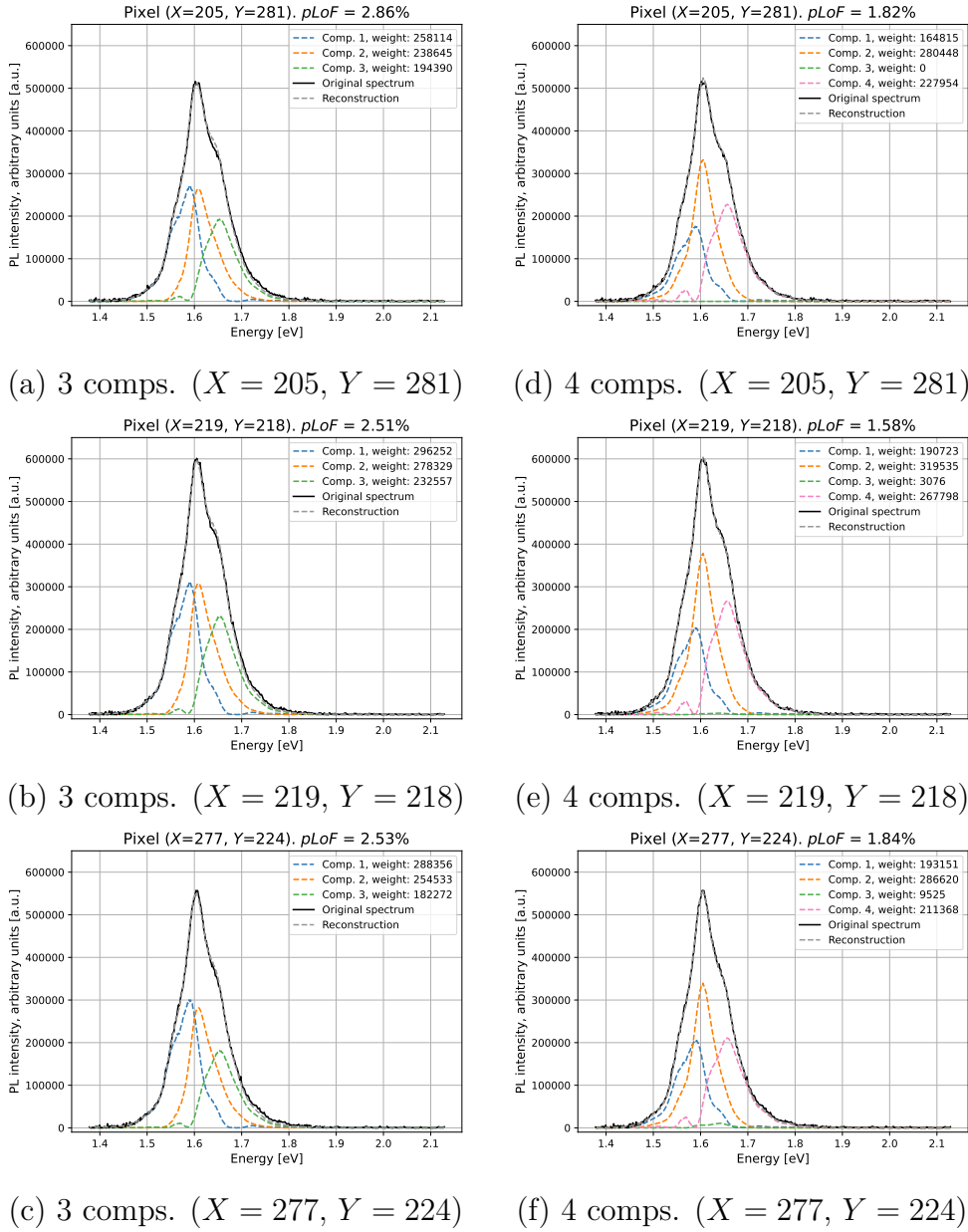


Figure 5.26: Comparison of how the models with three and four MCR components decompose and reconstruct selected pixel spectra from the circular region in Figure 5.25. Figures (a)-(c) represent the three-component model, and Figures (d)-(f) the model with four MCR components. The pure spectra are weighted by their respective concentrations, and the reconstructions are the sum of these weighted pure components.

Artifacts induced by the fabrication process can potentially explain why the three-component model has more problems with replicating the pixel spectra in the circular region than in other sample regions. The antisolvent may have

been dropped too rapidly onto the substrate during spin-coating so it washed the perovskite precursor solution off the middle of the substrate and left an area with a thinner perovskite layer behind [23]. Although it is unclear exactly how, this may have introduced a need for a fourth component to allow for small adjustments in the reconstructed spectra. However, more investigation must be carried out before anything can be concluded regarding how artifacts induced by the dropping of antisolvent affect the pixel spectra.

Regarding the need for a fourth MCR component, it is still unclear whether the extra component helps explain more of the underlying physics, or if its mere purpose is to minimize the reconstruction error. The score image of component 3 in the four-component MCR model (see Figure G.4 in Appendix G) shows that the concentration of this component is small in the circular region in the middle of the substrate. Adding a fourth component slightly improves the MCR model in terms of reconstruction error, as the fourth component seems to offer minor adjustments to the pixel spectra. However, Figure 5.23 suggests that the fourth component only improves the MCR model's reconstruction capacity in the circular region in the middle of the substrate and in the stripe at the bottom (where region C is). Since these regions may be consequences of various fabrication artifacts, it may be that components 1, 2, and 4 (corresponding to the three components in the three-component model) are most important for extracting relevant information from the images, as the two models perform equally well for a large majority of the pixels. The results from the four-component model will be investigated in the next section, but it will be compared to the simpler three-component model. The fourth component must be interpreted with caution.

## 5.9 Interpreting the MCR results (Sample 1)

### 5.9.1 Investigating the pure spectra of the MCR model with four components

Figure 5.27 shows the pure spectra for the MCR model with four components and  $\alpha = 0.15$ . Table 3 shows the locations (in eV) of these pure spectra's PL peaks. The table also shows which bromide-to-iodide ratios  $y$  these peaks correspond to, assuming that they are caused by BB processes and not defect-related SRH processes. The bromide-to-iodide ratios are calculated by using the linear regression model from Li et al. [32] presented in Equations (2.1) and (2.2).

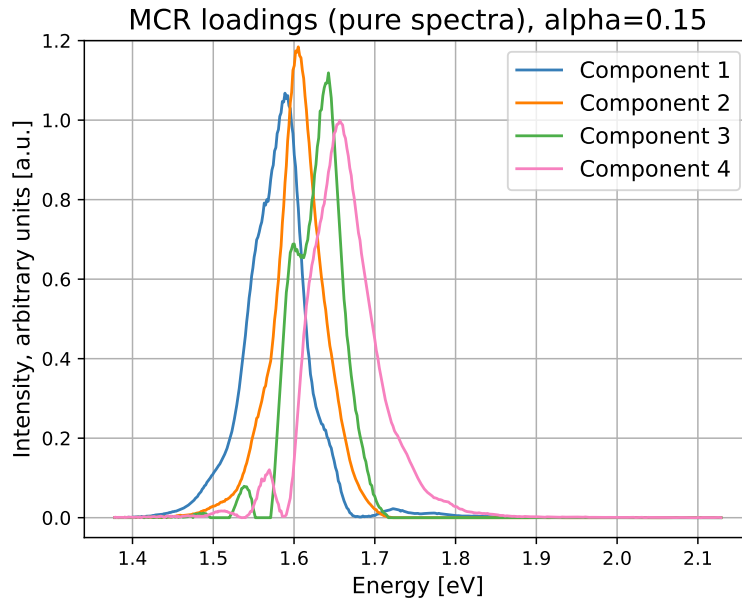


Figure 5.27: *Pure spectra for the MCR model with four components and  $\alpha = 0.15$ .*

Table 3: *Locations of the pure spectra's PL intensity peaks and the corresponding bromide-to-iodide ratios  $y$  for the regularized ( $\alpha = 0.15$ ) four-component MCR model. The halide ratios are predicted based on the linear regression model in Equations (2.1) and (2.2).*

	Peak energy [eV]	Predicted bromide-to-iodide ratio, $y$
Component 1	1.59	0.14
	1.73	0.34
Component 2	1.61	0.16
Component 3	1.49	0 <sup>a</sup>
	1.54	0.06
	1.60	0.15
	1.64	0.21
Component 4	1.51	0.01
	1.57	0.11
	1.66	0.24

<sup>a</sup>The model predicted a negative value of  $y$ . Since  $y = 0$  corresponds to 1.50 eV, the value was set to 0.

Table 3 shows that the pure components have several PL peaks that are potentially caused by phase segregation-related bandgap differences. Component 2's PL peak at 1.61 eV is consistent with emission from a mixed phase with approximately the same halide composition as the sample's overall halide stoichiometry (see Section 5.6). The remaining three components comprise PL peaks that may originate from phases with other halide compositions. The spectra of components 3 and 4 have some small peaks at around 1.49-1.54 eV that may originate from iodide-rich regions with  $y$  close to 0. Components 1, 3, and 4 have signals that may originate from regions with enhanced bromide fractions compared to the sample's overall stoichiometry. For instance, component 1 has a small PL peak at around 1.73 eV that may originate from a region with  $y = 0.34$ . Additionally, a PL intensity shoulder can be found at around 1.8 eV in component 4, which corresponds to  $y = 0.45$ . No PL signals are detected at the energies corresponding to bromide fractions that approach 1.



### 5.9.2 Comparison with the three-component model

For comparison, Table 4 shows the PL peaks and the corresponding bromide-to-iodide ratios  $y$  for the MCR model with three components and  $\alpha = 0$ . The locations of the peaks are almost identical to those of components 1, 2, and 4 in Table 3. A comparison of the spectra also reveals that these three components are very similar for the two models and that the major difference between the models is the addition of component 3 in the four-component model. Consequently, components 1, 2, and 4 in Table 3 seem more reliable than component 3. As discussed in Section 5.8.4, the additional component does not necessarily enhance the analysis since it only contributes to minor changes in the MCR model’s reconstructed spectra.

Table 4: *Locations of the pure spectra’s PL intensity peaks and the corresponding bromide-to-iodide ratios  $y$  for the non-regularized ( $\alpha = 0$ ) three-component model. The halide ratios are predicted based on the linear regression model in Equations (2.1) and (2.2).*

	Peak energy [eV]	Predicted bromide-to-iodide ratio, $y$
Component 1	1.59	0.13
	1.73	0.34
Component 2	1.61	0.16
Component 3	1.51	0.01
	1.57	0.11
	1.65	0.23

### 5.9.3 Discussion: segregated-phase stoichiometries reported in the literature

Most studies on light-induced phase segregation have investigated MAPb(Br <sub>$x$</sub> I <sub>$1-x$</sub> )<sub>3</sub> perovskite samples, and no known studies have reported the bromide-to-iodide ratios of the segregated phases in perovskites that are more similar to the samples studied in this work (i.e. with other A-site cations than only MA). Nonetheless, it is interesting to compare Table 3 with general phase segregation aspects reported for MAPb(Br <sub>$x$</sub> I <sub>$1-x$</sub> )<sub>3</sub>.

Hoke et al. [37] reported that MAPb(Br <sub>$x$</sub> I <sub>$1-x$</sub> )<sub>3</sub> segregates into lower-bandgap, iodide-rich minority domains and higher-bandgap, majority domains with a slightly increased bromide fraction. For instance, X-ray diffraction (XRD) measurements on MAPb(Br<sub>0.6</sub>I<sub>0.4</sub>)<sub>3</sub> thin films revealed that a minority phase

with an increased iodide concentration ( $x = 0.2$ ) and a majority phase with a slightly increased bromide concentration ( $x = 0.7$ ) formed under light-soaking [37]. This is consistent with the phase segregation mechanics that were described by Chen et al. [17] and explained in Section 2.2.3. For small iodide-rich domains to form, the bromide fraction must increase slightly elsewhere in the film.

PL signals that may originate from minority domains with significantly increased iodide concentrations were found in components 3 and 4 of the four-component MCR model. Components 3 and 4 also comprise PL signals at higher energies that may be attributed to BB emission from majority domains with a slightly enhanced bromide concentration (see Table 3). Hence, some of the signals in Table 3 seem to be consistent with the aforementioned observations from Hoke et al. [37].

Additionally, Hoke et al. [37] observed that the bromide-rich majority phase only had a *slightly* enhanced bromide concentration (i.e. they did not observe bromide-rich phases with bromide fractions approaching 1). This is consistent with the observations in this work since there are essentially no PL signals at energies above component 4's PL intensity shoulder at around 1.8 eV (see Figure 5.27), which corresponds to a bromide fraction of 0.45. Recall from Section 5.6 that a bromide fraction of 1 would have corresponded to a bandgap at 2.17 eV. That signals from regions with bromide fractions approaching 1 are not observed may be explained by the previously mentioned theory for light-induced phase segregation reported by Chen et al. [17]. Iodide-rich domains grow because photogenerated charge carriers accumulate in these domains. On the other hand, no such mechanism is responsible for creating bromide-rich regions. The higher bromide fractions in some regions of the sample are simply a consequence of the diffusion of bromide out of the iodide-rich domains. Therefore, one does not expect the bromide-rich regions to have as high bromide concentrations as the iodide concentrations of the iodide-rich domains.

The studies by Hoke et al. [37] and Bischak et al. [38] confirmed that the iodide-rich phase in  $\text{MAPb}(\text{Br}_x\text{I}_{1-x})_3$  ( $0.2 < x < 1$ ) perovskites was approximately  $\text{MAPb}(\text{Br}_{0.2}\text{I}_{0.8})_3$ , irrespective of the sample's overall, initial bromide-to-iodide ratio. However, some factors make direct comparisons with these observations challenging. First, the perovskites studied in the present work comprise other A-site cations than only MA. The literature suggests that adding other A-site cations makes the perovskites more stable under light exposure [68]. Second, the overall bromide-to-iodide ratio is 0.15, which is lower than

the range  $0.2 < x < 1$  for which Hoke et al. [37] observed the formation of the  $\text{MAPb}(\text{Br}_{0.2}\text{I}_{0.8})_3$  phase. Moreover, as mentioned in Section 5.2, previous studies have concluded that phase segregation mainly occurs in perovskites with large bromide fractions [4]. These two factors complicate direct comparison with the literature when it comes to which segregated phases to expect in the perovskites studied in this work. Even though the time series image in Section 5.2 suggests that phase segregation, or at least some kind of light soaking effect, occurs in the samples, no studies have been found that can help predict the stoichiometries of the phases that this work's perovskites will segregate into (in contrast to  $\text{MAPb}(\text{Br}_x\text{I}_{1-x})_3$ ). It is likely that at least some of the peaks in the pure spectra in Figure 5.27 originate from phases of different halide stoichiometries, and that the nature of these phases is similar to those in  $\text{MAPb}(\text{Br}_x\text{I}_{1-x})_3$ , but the challenge is to identify the signals that are caused by phase segregation and to distinguish them from signals caused by other mechanisms such as defects.

#### 5.9.4 Discussion: PL signals not directly related to phase segregation

The PL signals in Figure 5.27 may be caused by various mechanisms other than phase segregation. In multi-crystalline silicon, for instance, two-phonon processes create a PL knee at an energy slightly below the bandgap energy [21]. The PL spectra from perovskites often contain more than one peak, and Schötz et al. [69] state that the origin of the different PL features is debated. The perovskites studied by Schötz et al. [69] produced one dominant higher energy PL peak and a weaker peak or shoulder at a lower energy. They concluded that the higher energy peak was caused by emissions from near the surface and that the lower energy signal was caused by re-absorption effects in the bulk [69].

Schötz et al. [69] also presented some other explanations that have been proposed in the literature to explain the double PL peaks they observed. First, other researchers have used the coexistence of direct and indirect bandgap transitions caused by Rashba-splitting (see Section 2.2.2) to explain the double peak PL signals. However, Schötz et al. [69] showed that this explanation did not fit with the temperature dependence they observed for the PL signals. Second, defect-related recombination has been proposed as an explanation. As mentioned in Section 2.1.2, defect-related PL signals caused by SRH processes have energies lower than the bandgap energy. Moreover, the defects in perovskites are typically shallow and not deep-level defects [70, 71]. Third, the

coexistence of different crystalline phases has been used to explain the formation of more than one PL peak [69]. To summarize, the signals in Table 3 may have different origins and are not necessarily caused by phase segregation-related bandgap variations.

### 5.9.5 Discussion: correlated PL peaks

As was noted in Section 5.8.2, the pure spectra should ideally have been resolved into more components, as each of them usually comprises several PL peaks. One reason why the MCR algorithm does not resolve these signals into more components may be that some of the PL peaks are correlated in the sense that they are often present in the same pixels or sample regions. Take, for instance, component 3 in Table 3 as an example. The signal at 1.54 eV may represent an iodide-rich phase with the bromide-to-iodide ratio  $y = 0.06$  and the signal at 1.64 eV a bromide-rich phase with  $y = 0.21$ . For the iodide-rich regions to form, the bromide fraction has to increase in nearby regions, which may explain why these two peaks appear together.

Some MCR softwares offer employing unimodality constraints on the pure spectra [58], which could have helped resolve the multimodal spectra into more components. However, using these constraints was not tested in this work, because they are not implemented in the *pyMCR* package.

## 5.10 Pros and Cons Associated with PCA and MCR

### 5.10.1 PCA

A major advantage of PCA is its ease of use. First, it does not require any prior knowledge about the system. PCA simply identifies the directions of maximum variance in a given data set. Second, the number of components to use for analysis can be determined afterward by examining the variance explained by each principal component.

Another advantage of using PCA is that the method generally succeeds in reducing the dimensionality of the hyperspectral image. As was seen in Section 5.7, only the first few components contain useful information. The remaining components appear noisy and can hence be discarded. This ability to extract important information by constructing a new set of uncorrelated variables makes PCA ideal for obtaining an overview of the information in a hyperspectral image.

One disadvantage of PCA is that its objective (finding directions of maximum variance) is a mathematical rather than a physical criterion. This means that the method does not find the spectra of the system's underlying pure components, in contrast to MCR [57].

### 5.10.2 MCR

The main advantage of MCR compared to PCA is that in a situation where one assumes that the total signal is a superposition (linear combination) of a set of "pure" spectra, MCR can decompose this signal into its constituents. MCR estimates the pure spectra and the relative intensities of these components in each pixel.

One drawback of using MCR is that the number of components must be determined before running the algorithm. One simple approach tested in this work for determining the number of components was calculating "lack of fit" values for different regularization strengths. However, it is hard to conclude based on this method alone. Other alternatives include parallel analysis and cluster-aided MCR-ALS [72], but these methods were not tested in this work. Note that some software packages more advanced than *pyMCR* offer tools for estimating the number of components in the data set.

Another limitation of MCR is that the solutions to Equation (2.5) obtained by running the AR scheme are not unique (there are *ambiguities*). Therefore, there is no guarantee that the algorithm converges to the "true" solution of the system. The reason is that the algorithm stops optimization when the solution is "close enough" in terms of the given convergence criterion. This introduces uncertainty to the solution and makes it possible for several matrix pairs ( $\mathbf{C}$ ,  $\mathbf{S}^T$ ) to reconstruct the original data within this uncertainty [22]. However, applying constraints to the solutions helps generate more correct solutions [57].

An additional disadvantage associated with using MCR is that the algorithm must be provided with initial estimates of either the pure spectra or the concentrations. This requires some prior knowledge about the system. Moreover, since the solutions are not unique, it may be that different initializations lead to different solutions to the equation. Hence, it is not guaranteed that the initial guesses used to initialize the algorithm will yield the optimal ("true") solution. This work used spectral guesses, but guesses of the concentrations may also be used.

## 5.11 Limitations and Uncertainties

A major challenge in this work was to produce homogeneous perovskite samples that yielded satisfactory PL signals. As was noted in Section 5.3, the majority of samples produced rather weak PL signals. The fast degradation of perovskite samples may also have played a role here.

Moreover, the XRF imaging setup used in this work was not optimal for investigating the chemical composition of the perovskite layer, since this layer is very thin compared to the glass substrate. The setup is usually used to examine thicker samples. This may be the reason why the software's estimated concentrations seemed incorrect, as discussed in Section 5.4.

Another limitation is that the MCR software has many settings (regularization, convergence criteria) and an infinite number of ways to be initialized (using different initial spectra and concentrations). These choices influence the solution of Equation (2.5), but only a tiny subset of the different possibilities have been examined in this work. No other MCR softwares than *pyMCR* were tested in this work.

---

## 6 Conclusions

This work has revolved around developing data analysis techniques to map light-induced phase segregation in mixed-halide hybrid perovskite solar cells. A time-series image of a single line on one of the perovskite samples revealed that the line's PL spectrum blueshifted under light-soaking, which may be attributed to the formation of iodide-rich domains in other parts of the sample. However, other explanations for the blueshift can not be excluded, although they have not been identified yet. It remains unclear to which extent phase segregation will occur in the perovskite samples studied in this work, as phase segregation previously has been found to mainly occur in perovskites with larger bromide fractions.

The XRF images of the perovskite samples could not be used for identifying systematic variations in the halide concentrations across the sample's surface. Therefore, it was not possible to use these images to correlate PL signals at different energies and halide concentrations in an attempt to validate the analysis of the HSPL images. Based on the analysis of the hyperspectral EL image that was used to develop the data analysis techniques, larger regions of different bandgaps, probably caused by phase segregation, were expected. This was not observed in the samples studied in this work. The literature suggests that the iodide-rich domains are small compared to the resolutions of this work's XRF and HSPL setups.

The HSPL images showed considerable background noise from the Specim camera's detector array. Applying a background correction scheme that takes detector variations into account helped limit the influence of this noise. On the contrary, the scheme for normalizing the areas underneath the pixel spectra did not seem to enhance the analysis. When this scheme was not applied to the image, the first principal component captured the intensity variations anyway. The subsequent principal components were very similar whether the normalization was performed or not. In both cases, two of the principal components captured information that may be relevant for mapping phase segregation, but this lacks validation.

Regarding MCR, it proved difficult to build a model capable of resolving all the spectra's peaks and shoulders into separate components. Regularized regression gives the user more control over the model's components and offers a way to prevent the model from overfitting by adapting to noise. However, the challenge is to find the "true" components of the system, as more than one pair of matrices may solve the equations within the uncertainty of the convergence

criteria. In this work, the models with three or four components seemed to offer the best descriptions of the system.

The MCR spectra comprise several peaks and shoulders that may be attributed to the existence of phases of different halide stoichiometries. However, the literature suggests that also other factors, like re-absorption effects, may be responsible for these signals. A thorough investigation of the various PL signals and their origins must be conducted before one can unambiguously claim that one or more of them originate from light-induced phase segregation. However, the experimental setup and data analysis techniques investigated in this work seem promising for studying PL signals potentially caused by light-induced phase segregation. It is believed that HSPL imaging and the accompanying data analysis techniques may be capable of predicting regions of higher and lower concentrations of the halide species. PCA is suitable for conducting an exploratory analysis of the PL signals and MCR for decomposing the pixel spectra in the HSPL images.

### 6.1 Outlook and Further Work

In addition to obtaining an overview of the various PL signals present in the rather complex PL spectra of mixed-halide perovskites, a way to validate the analysis of the HSPL images must be in place to facilitate further advancements. Some chemical characterization methods that may be tested in combination with HSPL are X-ray diffraction (XRD), X-ray photoelectron spectroscopy (XPS), scanning electron microscopy with energy-dispersive X-ray spectroscopy (SEM/EDS), and secondary ion mass spectrometry (SIMS). Higher resolution HSPL and validation imaging setups may also help answer some of the questions about the sizes of the segregated domains raised in this work.

The HSPL imaging and accompanying data analysis techniques should be tested on perovskites of other chemistries than those studied here to assess the applicability of the methods described in this work. For instance, one may fabricate  $\text{MAPb}(\text{Br}_x\text{I}_{1-x})_3$  samples with  $x$  in the range where Hoke et al. [37] observed phase segregation to occur. As was discussed in Section 5.9, the iodide-rich phase in these samples is known to have  $x = 0.2$ . An interesting study would have been to investigate whether a PL signal corresponding to the bandgap of this phase can be found in the HSPL images of such samples and if this signal is detected by the PCA and MCR models.



## References

- [1] S. Wolf, J. Teitge, J. Mielke, F. Schütze, and C. Jaeger, “The European Green Deal — More Than Climate Neutrality,” *Intereconomics*, vol. 56, no. 2, pp. 99–107, Mar. 2021. DOI: [10.1007/s10272-021-0963-z](https://doi.org/10.1007/s10272-021-0963-z).
- [2] I. Kougias, N. Taylor, G. Kakoulaki, and A. Jäger-Waldau, “The role of photovoltaics for the European Green Deal and the recovery plan,” *Renewable and Sustainable Energy Reviews*, vol. 144, Jul. 2021. DOI: [10.1016/j.rser.2021.111017](https://doi.org/10.1016/j.rser.2021.111017).
- [3] International Energy Agency, “Renewables 2023. Analysis and forecasts to 2028,” Paris, France, Tech. Rep., Jan. 2024. [Online]. Available: [www.iea.org/reports/renewables-2023](http://www.iea.org/reports/renewables-2023) (visited on 05/21/2024).
- [4] L. Duan *et al.*, “Stability challenges for the commercialization of perovskite–silicon tandem solar cells,” *Nature Reviews Materials*, vol. 8, no. 4, pp. 261–281, Jan. 2023. DOI: [10.1038/s41578-022-00521-1](https://doi.org/10.1038/s41578-022-00521-1).
- [5] Fraunhofer Institute for Solar Energy Systems, ISE, “Photovoltaics Report,” Freiburg, Germany, Tech. Rep., May 2024. [Online]. Available: [www.ise.fraunhofer.de/en/publications/studies/photovoltaics-report.html](http://www.ise.fraunhofer.de/en/publications/studies/photovoltaics-report.html) (visited on 06/10/2024).
- [6] A. R. Zanatta, “The Shockley–Queisser limit and the conversion efficiency of silicon-based solar cells,” *Results in Optics*, vol. 9, Dec. 2022. DOI: [10.1016/j.rio.2022.100320](https://doi.org/10.1016/j.rio.2022.100320).
- [7] B. Ehrler, E. Alarcón-Lladó, S. W. Tabernig, T. Veeken, E. C. Garnett, and A. Polman, “Photovoltaics Reaching for the Shockley–Queisser Limit,” *ACS Energy Letters*, vol. 5, no. 9, pp. 3029–3033, Sep. 2020. DOI: [10.1021/acsenergylett.0c01790](https://doi.org/10.1021/acsenergylett.0c01790).
- [8] A. H. M. Smets, K. Jäger, O. Isabella, R. v. Swaaij, and M. Zeman, *Solar energy: the physics and engineering of photovoltaic conversion, technologies and systems*, 1st ed. Cambridge, England: UIT Cambridge Limited, 2016, ISBN: 978-1-906860-32-5.
- [9] A. Richter, M. Hermle, and S. W. Glunz, “Reassessment of the Limiting Efficiency for Crystalline Silicon Solar Cells,” *IEEE Journal of Photovoltaics*, vol. 3, no. 4, pp. 1184–1191, Sep. 2013. DOI: [10.1109/JPHOTOV.2013.2270351](https://doi.org/10.1109/JPHOTOV.2013.2270351).
- [10] National Renewable Energy Laboratory (NREL), *Interactive Best Research-Cell Efficiency Chart*. [Online]. Available: [www.nrel.gov/pv/interactive-cell-efficiency.html](http://www.nrel.gov/pv/interactive-cell-efficiency.html) (visited on 06/10/2024).
- [11] M. Grätzel, “The light and shade of perovskite solar cells,” *Nature Materials*, vol. 13, no. 9, pp. 838–842, Sep. 2014. DOI: [10.1038/nmat4065](https://doi.org/10.1038/nmat4065).

## REFERENCES

---

- [12] J. Hidalgo, A.-F. Castro-Méndez, and J.-P. Correa-Baena, “Imaging and Mapping Characterization Tools for Perovskite Solar Cells,” *Advanced Energy Materials*, vol. 9, no. 30, Aug. 2019. DOI: [10.1002/aenm.201904444](https://doi.org/10.1002/aenm.201904444).
- [13] C. Møller, “Crystal Structure and Photoconductivity of Cæsium Plumbahalides,” *Nature*, vol. 182, no. 4647, p. 1436, Nov. 1958. DOI: [10.1038/1821436a0](https://doi.org/10.1038/1821436a0).
- [14] D. Weber, “CH<sub>3</sub>NH<sub>3</sub>PbX<sub>3</sub>, ein Pb(II)-System mit kubischer Perowskitstruktur / CH<sub>3</sub>NH<sub>3</sub>PbX<sub>3</sub>, a Pb(II)-System with Cubic Perovskite Structure,” *Zeitschrift für Naturforschung B*, vol. 33, no. 12, pp. 1443–1445, Dec. 1978. DOI: [10.1515/znb-1978-1214](https://doi.org/10.1515/znb-1978-1214).
- [15] A. Kojima, K. Teshima, Y. Shirai, and T. Miyasaka, “Organometal Halide Perovskites as Visible-Light Sensitizers for Photovoltaic Cells,” *Journal of the American Chemical Society*, vol. 131, no. 17, pp. 6050–6051, May 2009. DOI: [10.1021/ja809598r](https://doi.org/10.1021/ja809598r).
- [16] M. Liu, M. B. Johnston, and H. J. Snaith, “Efficient planar heterojunction perovskite solar cells by vapour deposition,” *Nature*, vol. 501, no. 7467, pp. 395–398, Sep. 2013. DOI: [10.1038/nature12509](https://doi.org/10.1038/nature12509).
- [17] Z. Chen, G. Brocks, S. Tao, and P. A. Bobbert, “Unified theory for light-induced halide segregation in mixed halide perovskites,” *Nature Communications*, vol. 12, no. 1, May 2021. DOI: [10.1038/s41467-021-23008-z](https://doi.org/10.1038/s41467-021-23008-z).
- [18] K. Nguyen *et al.*, “Correlative imaging of optoelectronic properties for perovskite solar cells via hyperspectral luminescence imaging,” *Cell Reports Physical Science*, vol. 4, no. 10, Oct. 2023. DOI: [10.1016/j.xcrp.2023.101585](https://doi.org/10.1016/j.xcrp.2023.101585).
- [19] M. Xiao *et al.*, “A Fast Deposition-Crystallization Procedure for Highly Efficient Lead Iodide Perovskite Thin-Film Solar Cells,” *Angewandte Chemie International Edition*, vol. 53, no. 37, pp. 9898–9903, Jul. 2014. DOI: [10.1002/anie.201405334](https://doi.org/10.1002/anie.201405334).
- [20] F. Huang *et al.*, “Gas-assisted preparation of lead iodide perovskite films consisting of a monolayer of single crystalline grains for high efficiency planar solar cells,” *Nano Energy*, vol. 10, pp. 10–18, Nov. 2014. DOI: [10.1016/j.nanoen.2014.08.015](https://doi.org/10.1016/j.nanoen.2014.08.015).
- [21] T. Mehl, “Hyperspectral Photoluminescence Imaging of Silicon Wafers and Solar Cells,” Doctoral thesis, Norwegian University of Life Sciences, Ås, 2018.
- [22] G. M. Wyller, “Correlation of Defect Luminescence and Recombination in Multicrystalline Silicon,” M.S. thesis, Norwegian University of Life Sciences, Ås, 2015.

## REFERENCES

---

- [23] M. Saliba *et al.*, “How to Make over 20% Efficient Perovskite Solar Cells in Regular (  $n-i-p$  ) and Inverted (  $p-i-n$  ) Architectures,” *Chemistry of Materials*, vol. 30, no. 13, pp. 4193–4201, Jul. 2018. DOI: [10.1021/acs.chemmater.8b00136](https://doi.org/10.1021/acs.chemmater.8b00136).
- [24] D. J. J. Tay *et al.*, “Enabling a rapid SnO<sub>2</sub> chemical bath deposition process for perovskite solar cells,” *Sustainable Energy & Fuels*, vol. 7, no. 5, pp. 1302–1310, Feb. 2023. DOI: [10.1039/D2SE01475K](https://doi.org/10.1039/D2SE01475K).
- [25] N. Torabi *et al.*, “Progress and challenges in perovskite photovoltaics from single- to multi-junction cells,” *Materials Today Energy*, vol. 12, pp. 70–94, Jun. 2019. DOI: [10.1016/j.mtener.2018.12.009](https://doi.org/10.1016/j.mtener.2018.12.009).
- [26] W. Shockley and W. T. Read, “Statistics of the Recombinations of Holes and Electrons,” *Physical Review*, vol. 87, no. 5, pp. 835–842, Jan. 1952.
- [27] M. M. Ackermans, “Direct or Indirect? Unraveling the bandgap nature of metal halide perovskites,” M.S. thesis, TU Delft, Delft, Aug. 2020.
- [28] V. Sarritzu *et al.*, “Direct or Indirect Bandgap in Hybrid Lead Halide Perovskites?” *Advanced Optical Materials*, vol. 6, no. 10, Mar. 2018. DOI: [10.1002/adom.201701254](https://doi.org/10.1002/adom.201701254).
- [29] T. Wang *et al.*, “Indirect to direct bandgap transition in methylammonium lead halide perovskite,” *Energy & Environmental Science*, vol. 10, no. 2, pp. 509–515, Feb. 2017. DOI: [10.1039/C6EE03474H](https://doi.org/10.1039/C6EE03474H).
- [30] T. Wang, B. Daiber, and B. Ehrler, *Indirect to direct bandgap transition in methylammonium lead halide perovskite*, Advances in Engineering. [Online]. Available: [www.advanceseng.com/indirect-to-direct-bandgap-transition-methylammonium-lead-halide-perovskite/](http://www.advanceseng.com/indirect-to-direct-bandgap-transition-methylammonium-lead-halide-perovskite/) (visited on 01/24/2024).
- [31] S. A. Kulkarni, T. Baikie, P. P. Boix, N. Yantara, N. Mathews, and S. Mhaisalkar, “Band-gap tuning of lead halide perovskites using a sequential deposition process,” *Journal of Materials Chemistry A*, vol. 2, no. 24, pp. 9221–9225, May 2014. DOI: [10.1039/C4TA00435C](https://doi.org/10.1039/C4TA00435C).
- [32] Y. Li *et al.*, “Bandgap tuning strategy by cations and halide ions of lead halide perovskites learned from machine learning,” *RSC Advances*, vol. 11, no. 26, pp. 15 688–15 694, Apr. 2021. DOI: [10.1039/D1RA03117A](https://doi.org/10.1039/D1RA03117A).
- [33] R. Wang, T. Huang, J. Xue, J. Tong, K. Zhu, and Y. Yang, “Prospects for metal halide perovskite-based tandem solar cells,” *Nature Photonics*, vol. 15, no. 6, pp. 411–425, Jun. 2021. DOI: [10.1038/s41566-021-00809-8](https://doi.org/10.1038/s41566-021-00809-8).
- [34] R. G. Balakrishna, S. M. Kobosko, and P. V. Kamat, “Mixed Halide Perovskite Solar Cells. Consequence of Iodide Treatment on Phase Seg-

## REFERENCES

---

- regation Recovery,” *ACS Energy Letters*, vol. 3, no. 9, pp. 2267–2272, Sep. 2018. DOI: [10.1021/acsenergylett.8b01450](https://doi.org/10.1021/acsenergylett.8b01450).
- [35] D. J. Slotcavage, H. I. Karunadasa, and M. D. McGehee, “Light-Induced Phase Segregation in Halide-Perovskite Absorbers,” *ACS Energy Letters*, vol. 1, no. 6, pp. 1199–1205, Dec. 2016. DOI: [10.1021/acsenergylett.6b00495](https://doi.org/10.1021/acsenergylett.6b00495).
- [36] J. H. Noh, S. H. Im, J. H. Heo, T. N. Mandal, and S. I. Seok, “Chemical Management for Colorful, Efficient, and Stable Inorganic–Organic Hybrid Nanostructured Solar Cells,” *Nano Letters*, vol. 13, no. 4, pp. 1764–1769, Apr. 2013. DOI: [10.1021/nl400349b](https://doi.org/10.1021/nl400349b).
- [37] E. T. Hoke, D. J. Slotcavage, E. R. Dohner, A. R. Bowring, H. I. Karunadasa, and M. D. McGehee, “Reversible photo-induced trap formation in mixed-halide hybrid perovskites for photovoltaics,” *Chemical Science*, vol. 6, no. 1, pp. 613–617, Dec. 2014. DOI: [10.1039/C4SC03141E](https://doi.org/10.1039/C4SC03141E).
- [38] C. G. Bischak *et al.*, “Origin of Reversible Photoinduced Phase Separation in Hybrid Perovskites,” *Nano Letters*, vol. 17, no. 2, pp. 1028–1033, Feb. 2017. DOI: [10.1021/acs.nanolett.6b04453](https://doi.org/10.1021/acs.nanolett.6b04453).
- [39] W. Mao *et al.*, “Light-induced reversal of ion segregation in mixed-halide perovskites,” *Nature Materials*, vol. 20, no. 1, pp. 55–61, Jan. 2021. DOI: [10.1038/s41563-020-00826-y](https://doi.org/10.1038/s41563-020-00826-y).
- [40] C. Yong-Liang *et al.*, “Progress in perovskite solar cells based on different buffer layer materials,” *Acta Physica Sinica*, vol. 69, no. 13, Jul. 2020. DOI: [10.7498/aps.69.20200543](https://doi.org/10.7498/aps.69.20200543).
- [41] A. Kahn, “Fermi level, work function and vacuum level,” *Materials Horizons*, vol. 3, no. 1, pp. 7–10, Jan. 2016. DOI: [10.1039/C5MH00160A](https://doi.org/10.1039/C5MH00160A).
- [42] S. Ravishankar *et al.*, “Influence of Charge Transport Layers on Open-Circuit Voltage and Hysteresis in Perovskite Solar Cells,” *Joule*, vol. 2, no. 4, pp. 788–798, Apr. 2018. DOI: [10.1016/j.joule.2018.02.013](https://doi.org/10.1016/j.joule.2018.02.013).
- [43] I. E. Høiaas, “Hyperspectral imaging as a tool to study solar induced photoluminescence from PV modules,” M.S. thesis, Norwegian University of Life Sciences, Ås, 2019.
- [44] K. N. Shinde, S. Dhoble, H. Swart, and K. Park, *Phosphate Phosphors for Solid-State Lighting* (Springer Series in Materials Science). Heidelberg, Germany: Springer Berlin, 2012, ISBN: 978-3-642-34312-4. DOI: [10.1007/978-3-642-34312-4](https://doi.org/10.1007/978-3-642-34312-4).
- [45] A. M. Soufiani *et al.*, “Implied Open-circuit Voltage Imaging via a Single Bandpass Filter Method—Its First Application in Perovskite Solar Cells,” *Advanced Functional Materials*, vol. 33, no. 3, Jan. 2023. DOI: [10.1002/adfm.202210592](https://doi.org/10.1002/adfm.202210592).

## REFERENCES

---

- [46] P. Würfel, S. Finkbeiner, and E. Daub, “Generalized Planck’s radiation law for luminescence via indirect transitions,” *Applied Physics A*, vol. 60, no. 1, pp. 67–70, Jan. 1995. DOI: [10.1007/BF01577615](https://doi.org/10.1007/BF01577615).
- [47] V. Lodhi, D. Chakravarty, and P. Mitra, “Hyperspectral Imaging for Earth Observation: Platforms and Instruments,” *Journal of the Indian Institute of Science*, vol. 98, no. 4, pp. 429–443, Dec. 2018. DOI: [10.1007/s41745-018-0070-8](https://doi.org/10.1007/s41745-018-0070-8).
- [48] J. M. Amigo, I. Martí, and A. Gowen, “Hyperspectral Imaging and Chemometrics: A Perfect Combination for the Analysis of Food Structure, Composition and Quality,” in *Data Handling in Science and Technology*, vol. 28, Amsterdam, The Netherlands: Elsevier, 2013, pp. 343–370, ISBN: 978-0-444-59528-7. DOI: [10.1016/B978-0-444-59528-7.00009-0](https://doi.org/10.1016/B978-0-444-59528-7.00009-0).
- [49] F. Dell’Endice, J. Nieke, B. Koetz, M. E. Schaepman, and K. Itten, “Improving radiometry of imaging spectrometers by using programmable spectral regions of interest,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 64, no. 6, pp. 632–639, Nov. 2009. DOI: [10.1016/j.isprsjprs.2009.05.007](https://doi.org/10.1016/j.isprsjprs.2009.05.007).
- [50] M. J. Pushie, I. J. Pickering, M. Korbas, M. J. Hackett, and G. N. George, “Elemental and Chemically Specific X-ray Fluorescence Imaging of Biological Systems,” *Chemical Reviews*, vol. 114, no. 17, pp. 8499–8541, Sep. 2014. DOI: [10.1021/cr4007297](https://doi.org/10.1021/cr4007297).
- [51] C. Ossig *et al.*, “Four-Fold Multi-Modal X-ray Microscopy Measurements of a Cu(In,Ga)Se<sub>2</sub> Solar Cell,” *Materials*, vol. 14, no. 1, p. 228, Jan. 2021. DOI: [10.3390/ma14010228](https://doi.org/10.3390/ma14010228).
- [52] Y. Luo *et al.*, “The Relationship between Chemical Flexibility and Nanoscale Charge Collection in Hybrid Halide Perovskites,” *Advanced Functional Materials*, vol. 28, no. 18, May 2018. DOI: [10.1002/adfm.201706995](https://doi.org/10.1002/adfm.201706995).
- [53] S. Raschka and V. Mirjalili, *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2* (Expert insight), 3rd ed. Birmingham, United Kingdom: Packt Publishing Limited, 2019, ISBN: 978-1-78995-575-0.
- [54] D. C. Lay, S. R. Lay, and J. J. McDonald, *Linear Algebra and Its Applications*, 5th ed. Harlow, United Kingdom: Pearson Education Limited, 2016, ISBN: 978-1-292-09223-2.
- [55] S. Kokot, M. Grigg, H. Panayiotou, and T. D. Phuong, “Data Interpretation by some Common Chemometrics Methods,” *Electroanalysis*, vol. 10,

## REFERENCES

---

- no. 16, pp. 1081–1088, Aug. 1998. DOI: [10.1002/\(SICI\)1521-4109\(199811\)10:16<1081::AID-ELAN1081>3.0.CO;2-X](https://doi.org/10.1002/(SICI)1521-4109(199811)10:16<1081::AID-ELAN1081>3.0.CO;2-X).
- [56] H. Abdi and L. J. Williams, “Principal component analysis,” *WIREs Computational Statistics*, vol. 2, no. 4, pp. 433–459, Jul. 2010. DOI: [10.1002/wics.101](https://doi.org/10.1002/wics.101).
- [57] X. Zhang and R. Tauler, “Application of Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) to remote sensing hyperspectral imaging,” *Analytica Chimica Acta*, vol. 762, pp. 25–38, Jan. 2013. DOI: [10.1016/j.aca.2012.11.043](https://doi.org/10.1016/j.aca.2012.11.043).
- [58] J. Jaumot, R. Gargallo, A. De Juan, and R. Tauler, “A graphical user-friendly interface for MCR-ALS: A new tool for multivariate curve resolution in MATLAB,” *Chemometrics and Intelligent Laboratory Systems*, vol. 76, no. 1, pp. 101–110, Mar. 2005. DOI: [10.1016/j.chemolab.2004.12.007](https://doi.org/10.1016/j.chemolab.2004.12.007).
- [59] C. H. Camp, “pyMCR: A Python Library for Multivariate Curve Resolution Analysis with Alternating Regression (MCR-AR),” *Journal of Research of the National Institute of Standards and Technology*, vol. 124, Jun. 2019. DOI: [10.6028/jres.124.018](https://doi.org/10.6028/jres.124.018).
- [60] M. O’Kane, *How to Make Efficient Perovskite Solar Cells in a Glove Box*, Ossila. [Online]. Available: [www.ossila.com/pages/how-to-make-perovskites-solar-cells-in-a-glovebox-part-2](http://www.ossila.com/pages/how-to-make-perovskites-solar-cells-in-a-glovebox-part-2) (visited on 04/22/2024).
- [61] W. Zhang *et al.*, “Organic Salts as p-Type Dopants for Efficient LiTFSI-Free Perovskite Solar Cells,” *ACS Applied Materials & Interfaces*, vol. 12, no. 30, pp. 33 751–33 758, Jul. 2020. DOI: [10.1021/acsami.0c08322](https://doi.org/10.1021/acsami.0c08322).
- [62] A. S. Flø, “Hyperspectral Imaging as a Tool for Characterization of Multicrystalline Silicon Wafers,” Doctoral thesis, Norwegian University of Life Sciences, Ås, 2014.
- [63] J. Mooney and P. Kambhampati, “Get the Basics Right: Jacobian Conversion of Wavelength and Energy Scales for Quantitative Analysis of Emission Spectra,” *The Journal of Physical Chemistry Letters*, vol. 4, no. 19, pp. 3316–3318, Oct. 2013. DOI: [10.1021/jz401508t](https://doi.org/10.1021/jz401508t).
- [64] J. Hass, C. Heil, and M. D. Weir, *Thomas’ Calculus in SI Units*, 14th ed. Harlow, United Kingdom: Pearson Education Limited, 2019, ISBN: 978-1-292-25322-0.
- [65] M. G. Larson and F. Bengzon, *The finite element method: theory, implementation, and applications* (Texts in computational science and engineering), 1st ed. Heidelberg, Germany: Springer Berlin, 2013, ISBN: 978-3-642-33286-9.

## REFERENCES

---

- [66] S. Siegrist, P. Nandi, R. K. Kothandaraman, A. Abdesslem, A. N. Tiwari, and F. Fu, “Understanding Coating Thickness and Uniformity of Blade-Coated SnO<sub>2</sub> Electron Transport Layer for Scalable Perovskite Solar Cells,” *Solar RRL*, vol. 7, no. 14, May 2023. DOI: [10.1002/solr.202300273](https://doi.org/10.1002/solr.202300273).
- [67] A. J. Barker *et al.*, “Defect-Assisted Photoinduced Halide Segregation in Mixed-Halide Perovskite Thin Films,” *ACS Energy Letters*, vol. 2, no. 6, pp. 1416–1424, Jun. 2017. DOI: [10.1021/acseenergylett.7b00282](https://doi.org/10.1021/acseenergylett.7b00282).
- [68] T. Duong *et al.*, “Light and Electrically Induced Phase Segregation and Its Impact on the Stability of Quadruple Cation High Bandgap Perovskite Solar Cells,” *ACS Applied Materials & Interfaces*, vol. 9, no. 32, pp. 26 859–26 866, Aug. 2017. DOI: [10.1021/acsami.7b06816](https://doi.org/10.1021/acsami.7b06816).
- [69] K. Schötz *et al.*, “Double peak emission in lead halide perovskites by self-absorption,” *Journal of Materials Chemistry C*, vol. 8, no. 7, pp. 2289–2300, Feb. 2020. DOI: [10.1039/C9TC06251C](https://doi.org/10.1039/C9TC06251C).
- [70] Y. Yuan *et al.*, “Shallow defects and variable photoluminescence decay times up to 280  $\mu$ s in triple-cation perovskites,” *Nature Materials*, vol. 23, no. 3, pp. 391–397, Mar. 2024. DOI: [10.1038/s41563-023-01771-2](https://doi.org/10.1038/s41563-023-01771-2).
- [71] Y. Yamada *et al.*, “Dynamic Optical Properties of CH<sub>3</sub>NH<sub>3</sub>PbI<sub>3</sub> Single Crystals As Revealed by One- and Two-Photon Excited Photoluminescence Measurements,” *Journal of the American Chemical Society*, vol. 137, no. 33, pp. 10 456–10 459, Aug. 2015. DOI: [10.1021/jacs.5b04503](https://doi.org/10.1021/jacs.5b04503).
- [72] H. Motegi *et al.*, “Identification of Reliable Components in Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS): A Data-Driven Approach across Metabolic Processes,” *Scientific Reports*, vol. 5, no. 1, Nov. 2015. DOI: [10.1038/srep15710](https://doi.org/10.1038/srep15710).
- [73] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, Oct. 2011.

---

## Appendices

### A Python code used for data preprocessing

#### A.1 Reading the hyperspectral EL image

The following box shows the function used to open the hyperspectral EL image file.

```
import numpy as np
import h5py
import os

def load_h5(filename, dirname=None):
    """
    This function takes a filename (and directory name) as
        input,
    and returns a dictionary containing the hypercube
    and wavelengths as outputs.

    If dirname=None, the function assumes that the file given
        by
    'filename' is located in the current working directory.
    """

    if dirname:
        dir_file=os.path.join(dirname, filename)
    else:
        dir_file=os.path.join(os.getcwd(), filename)

    # Opening the file, saving the data as an array
    # and the wavelength region
    f = h5py.File(dir_file, 'r')
    cube = np.array(f['Cube']['Images'])
    wavelength = np.array(f['Cube']['Wavelength'])
    f.close()

    return {'cube': cube, 'wavelength': wavelength}
```

The data is loaded into Python by using the following lines of code:

```
data = load_h5(filename='PSC_40mA_LightSoak60s_Exposure1s_
Start580nm_Stop880nm_Step10nm.h5')
cube, wavelength = data["cube"], data["wavelength"]
cube = np.transpose(cube, (1, 2, 0))
wl = [float(wl) for wl in wavelength]
```



## A.2 Reading the hyperspectral PL images (Specim)

---

The third line makes sure that the hypercube has the wavelength bands as its third dimension.

Rotation and cropping are performed using these lines of code:

```
from scipy import ndimage

def rotation(image, angle):
    """
    Rotating the image by an angle of 'angle' degrees.
    """
    return ndimage.rotate(image[:, :, :], angle, axes=(0,1))

cube_rot = rotation(cube, angle=-5)
hypercube=cube_rot[200:1080,250:1080,:]
```

Negative values are removed this way:

```
image[image < 0] = 0
```

## A.2 Reading the hyperspectral PL images (Specim)

The following code was used to load a hyperspectral PL images recorded by the Specim camera, as well as the accompanying header file, from a folder.

```
import os
from spectral import *

def load_hyperspectral(dirname):
    dir_path=os.path.join(os.getcwd(), dirname)
    files = os.listdir(dir_path)

    for file in files:
        if file.endswith('.raw'):
            hyperspectral_image = os.path.join(dirname, file)
        elif file.endswith('.hdr'):
            hdr_file = os.path.join(dirname, file)

    img = envi.open(hdr_file, hyperspectral_image)
    hyperim = img.load()[:, :, :]
    header = envi.read_envi_header(hdr_file)

    return hyperim, header

# Image and header
im, header = load_hyperspectral("Hyperspectral")

# Wavelength bands
```

### A.3 Jacobian conversion from wavelength (nm) to energy (eV) units

---

```
wl = header["wavelength"]
wl = [float(w) for w in wl]
```

#### A.2.1 Background noise correction

The background noise correction is performed using these lines of code:

```
# The first dark frame:
autodarkline = int(header["autodarkstartline"])

# Construct background noise matrix (median of 50 last frames):
BNM = np.median(im[autodarkline+10:, :, :], axis=0)

# Perform the background noise correction
for i in range(im.shape[0]):
    im[i, :, :] = im[i, :, :] - BNM
im = im.astype('float64')
```

### A.3 Jacobian conversion from wavelength (nm) to energy (eV) units

The following function is an implementation of the Jacobian transformation, converting intensity values from a wavelength to an energy scale:

```
def convert_to_eV(image, wl):
    """
    Takes a hyperspectral image as input (shape: [y,x,lambda]),
    in addition to a list containing wavelengths
    in nm representing the wavelength bands.

    Converts the unit from nm to eV by using the Jacobian
    transformation.

    Returns a list containing the energy values corresponding
    to
    the wavelength bands, and the converted image.

    The transformation is based on the following article:
    https://pubs.acs.org/doi/10.1021/jz401508t
    """

    # Energy bands
    eV = [float(1239.8/float(w)) for w in wl]

    image_ev = np.zeros(image.shape)
```

#### A.4 Scaling the intensity values by normalizing the areas underneath the pixel spectra

---

```
# Iterate through the wavelength bands
for i in range(image.shape[2]):
    # Jacobian transformation
    image_ev[:, :, i] = image[:, :, i]*1239.9/(eV[i]**2)
return eV, image_ev

# Example of how to use this function:
eV, image_ev = convert_to_eV(hypercube, wl)
```

#### A.4 Scaling the intensity values by normalizing the areas underneath the pixel spectra

This function is an implementation of the trapezoidal rule for numerical integration:

```
def trapezoidal(x, y):
    """
    Numerical integration using the Trapezoidal rule.
    x represents the independent variable, and y represents
        function values.

    Important: The lengths of x and y must be equal.
    """
    I = 0
    h = x[1] - x[0] # Interval width

    # Add the endpoints
    I += (y[0]+y[-1])

    # Add interior points with correct weight
    I += 2 * y[1:-1].sum()

    # Take interval width into consideration
    I *= h / 2

    return I
```

To apply the trapezoidal rule, the interval widths need to be uniform. Therefore, if the spacing between the wavelength bands is non-uniform, interpolation is required to construct intervals of uniform widths. The following function uses piecewise linear interpolation (i.e. it assumes line segments between the spectrum's values) and integrates this signal using the trapezoidal rule:

```
def integrate_trapezoidal(wl, pl, num_intervals=1000):
```

#### A.4 Scaling the intensity values by normalizing the areas underneath the pixel spectra

---

```
xnew = np.linspace(np.min(wl), np.max(wl), num_intervals+1)
ynew = np.interp(xnew, wl, pl)

return trapezoidal(xnew, ynew)
```

The following function performs the normalization:

```
def normalize_total_intensity(wl, image):
    """
    This function uses the trapezoidal rule to integrate each
        pixel spectrum.
    The spectra are then divided by the areas underneath them
        to
    yield spectra with unit area.

    Input: list of wavelengths ('wl') and unscaled image ('
        image')
    Returns: area-normalized image ('image_norm') and an image
        containing the areas
        underneath the spectra ('image_intensities')
    """
    image_norm = np.zeros(image.shape)
    image_intensities = np.zeros((image.shape[0], image.shape[1]
        ))

    for i in range(image.shape[0]):
        for j in range(image.shape[1]):
            spectrum = image[i,j,:] # Extract pixel spectrum
            A = integrate_trapezoidal(wl, spectrum) #
                Normalization
                constant (equal to
                the area under the
                curve)

            image_norm[i,j,:] = spectrum / A
            image_intensities[i,j] = A
    return image_norm, image_intensities

# Example of how to use the function:
image_nm_norm, intensities = normalize_total_intensity(wl,
    image_nm)
```

---

## B Python code used for data analysis

This section contains Python code used to run PCA and MCR.

### B.1 PCA

PCA is performed using the *Scikit-learn* package [73], and the following lines of code:

```
from sklearn.decomposition import PCA

# Reshape the hyperspectral data into a data matrix
X = image_ev_norm_cropped.reshape(
    image_ev_norm_cropped.shape[0]*image_ev_norm_cropped.shape[1],
    image_ev_norm_cropped.shape[2])

# Perform PCA
pca = PCA(n_components=100) # Pick out the first 100 PC's
score_matrix = pca.fit_transform(X) # T
loading_matrix = pca.components_ # P^T
ex_var = pca.explained_variance_ratio_
```

Here, 'image\_ev\_norm\_cropped' is a hyperspectral image in which the areas underneath the pixel spectra are normalized and the scale is converted from wavelength to energy units. Moreover, the image is cropped so that the edge of the perovskite film is removed.

When performing PCA, it was observed that the score images returned by *scikit-learn* had a mean of zero. The reason for this may be that *scikit-learn* centers the features (corresponding to the wavelength bands) before performing the singular value decomposition used to calculate the principal components (see the documentation of *scikit-learn* version 1.2.2). A pixel's score value for a given principal component is therefore different from the linear combination of its intensity values with the principal component's loadings as weights. For easier interpretation of the score values, it was therefore decided to calculate the scores manually as linear combinations of the original pixel values, using the loadings as weights. This is done by computing the product  $\mathbf{T} = \mathbf{D}\mathbf{P}$ , where  $\mathbf{T}$  is the score matrix,  $\mathbf{D}$  is the original data matrix, and  $\mathbf{P}$  is the transpose of the loading matrix  $\mathbf{P}^T$ .

## B.1 PCA

---

The PCA score images are obtained using this line of code:

```
img_pc = np.dot(X, pca.components_.T).reshape((*  
                                              image_ev_norm_cropped.shape[:2]  
                                              , 100))
```

Loadings are obtained this way:

```
loadings = loading_matrix.T
```

## B.2 MCR

The *pyMCR* package [59] is used for performing MCR on the hyperspectral images. Initial estimates of the pure spectra are used to initialize the MCR algorithm. The function below generates  $N$  evenly spaced, Gaussian-shaped spectra. The leftmost and rightmost spectra are created so that their peaks coincide with the extremes of the PC1 or PC2 loadings (depending on whether the image is normalized or not), obtained from a preceding PCA step. If  $N$  is greater than two, the remaining spectra are placed between the two aforementioned spectra, equidistant from each other.

```
def initialize_spectra(n_components, pca_loadings, pc_init):
    """
    This function generates, based on the peaks in the PC1
        loading plot,
    evenly spaced, Gaussian-shaped spectra.
    These spectra are later used to initialize the MCR
        algorithm.

    Inputs:
        - n_components : Number of spectra to generate
        - pca_loadings : Loadings from PCA

    Output:
        - spectra : a numpy array containing the spectra.
            Shape: (n_components, number of wavelength
                    bands)

    """
    # Locate peaks in PC1 or PC2 loading
    if pc_init == 1:
        pc1_loading = pca_loadings[:,0]
        peak1 = eV[np.argmax(pc1_loading)]
        peak2 = eV[np.argmin(pc1_loading)]
    elif pc_init == 2:
        pc2_loading = pca_loadings[:,1]
        peak1 = eV[np.argmax(pc2_loading)]
        peak2 = eV[np.argmin(pc2_loading)]

    # Calculate the distance between successive spectra
    increment = np.abs(peak2 - peak1) / (n_components - 1)

    # Generate evenly spaced, Gaussian-shaped spectra
    spectra = []
    for i in range(n_components):
        sp = np.exp(-(np.array(eV)-min(peak1, peak2) - i *
                                   increment)**2/(2*0.03**
                                   2))
```

```

        spectra.append(sp)

    spectra = np.vstack(tuple(spectra))
    return spectra

# The function is used like this:
n_components = 5
initial_spectra = initialize_spectra(n_components=n_components,
                                    pca_loadings=loadings)

```

Running the MCR algorithm is implemented as follows:

```

from pymcr.mcr import McrAR

hyperim = image_ev_norm_cropped # Same image as PCA code
                                example

# Note that the unfolding has to be implemented manually
# when using the pyMCR package
D = hyperim.reshape((hyperim.shape[0]*hyperim.shape[1], hyperim
                    .shape[2]))

# Running MCR (with Ridge Regression)
mcrar = McrAR(tol_increase=5.0, c_regr=Ridge(alpha=0.05),
              st_regr=Ridge(alpha=0.05),
              c_constraints=[ConstraintNonneg
                             ()], st_constraints=[
                             ConstraintNonneg()])
mcrar.fit(D, ST=initial_spectra)

```

The scores (concentrations) and loadings (pure spectra) are obtained using the following lines of code:

```

# Concentrations:
C = mcrar.C_opt_
scores = C.reshape(*hyperim.shape[:2], n_components)

# Pure spectra:
ST = mcrar.ST_opt_

```

The errors ( $MSE$ ) after each half-iteration can be accessed through the 'err' attribute:

```

# Errors (MSE):
error = mcrar.err

```

The reconstructions of the pixel spectra are obtained by multiplying the score matrix and the loading matrix:



## B.2 MCR

---

```
R = np.dot(C,ST) # Matrix, shape=(num_pixels,num_bands)
reconstructed = R.reshape(*hyperim.shape) # 3D array, shape=(y,
                                         x,num_bands)
```

Residual images:

```
E = D - R # Calculate residual matrix
E_images = E.reshape(*hyperim.shape) # Residual images (one
                                     per band)
```

---

## C Results obtained by analyzing the hyperspectral EL image

### C.1 Preprocessing

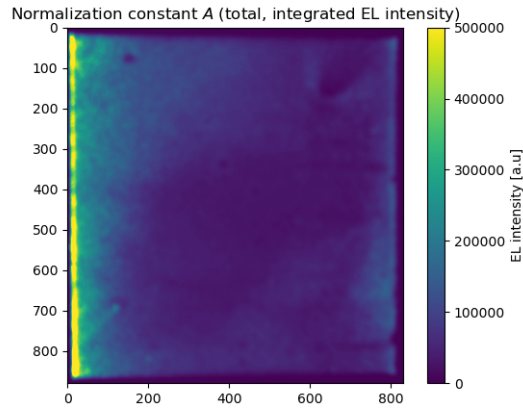


Figure C.1: *Normalization constants calculated by numerical integration. These values give the total EL intensities emitted by each pixel.*

### C.2 Pipeline 1: PCA

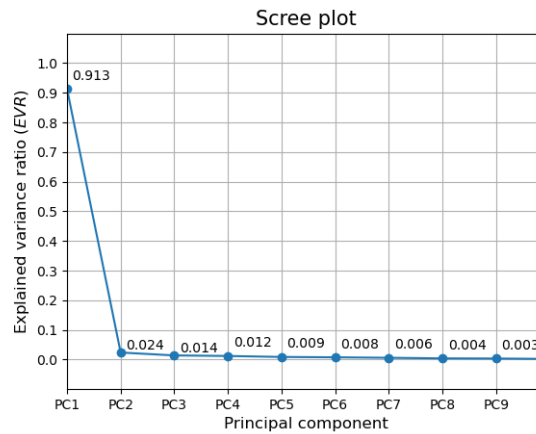


Figure C.2: *Scree plot showing the fraction of the total variance explained by each principal component. Only the first nine principal components are included in the plot.*

## C.2 Pipeline 1: PCA

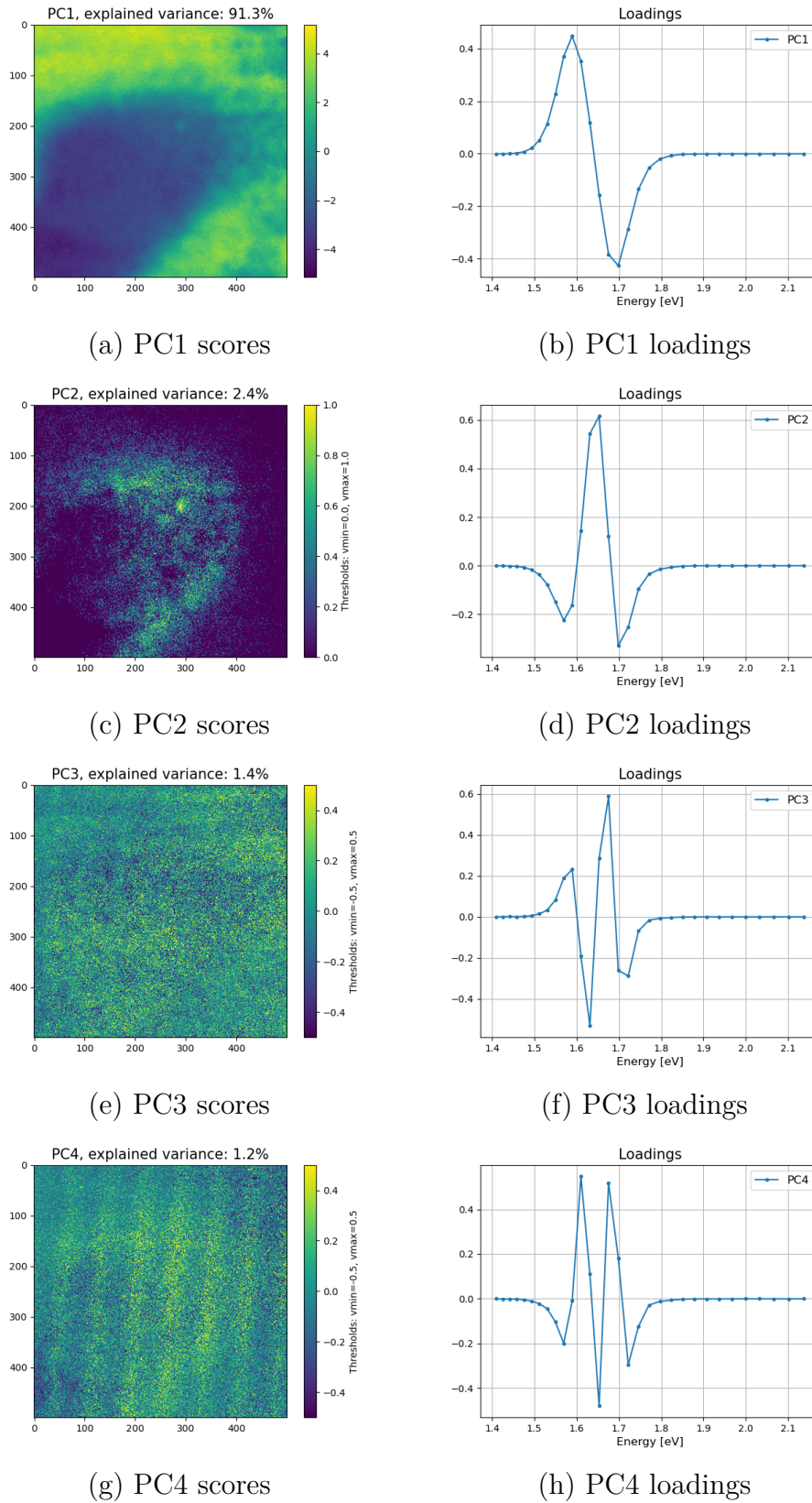
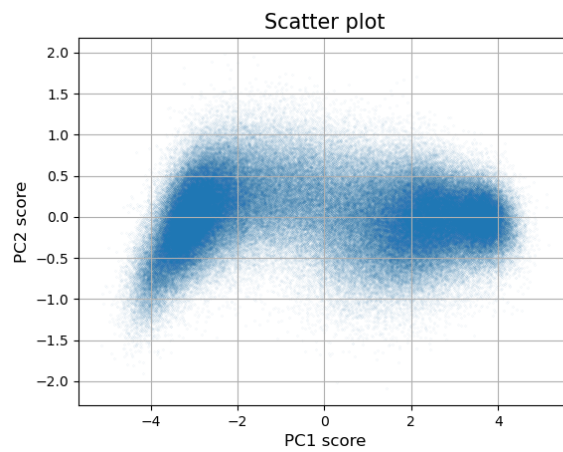
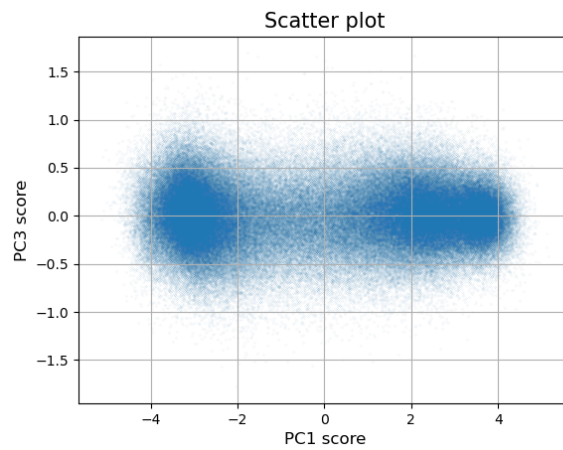


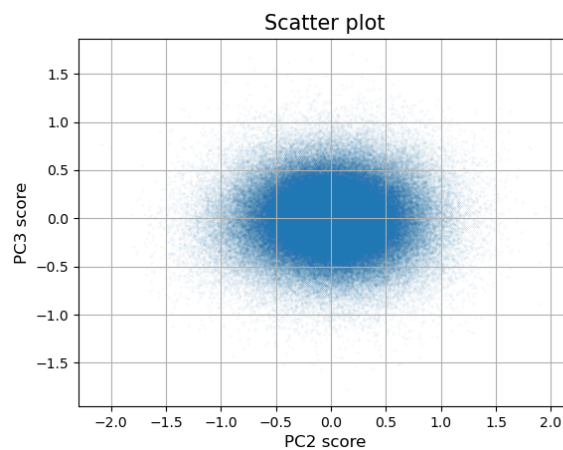
Figure C.3: Scores and loadings of the first four principal components. Note that the colorbar values are thresholded in Figures (c), (e), and (g), for visual clarity.



(a) PC1 vs. PC2



(b) PC1 vs. PC3



(c) PC2 vs. PC3

Figure C.4: *Scatter plots visualizing the pixels in lower-dimensional spaces spanned by the first three principal component axes.*

## C.3 Pipeline 2: PCA + MCR

### C.3.1 Using 2 MCR components (without regularization)

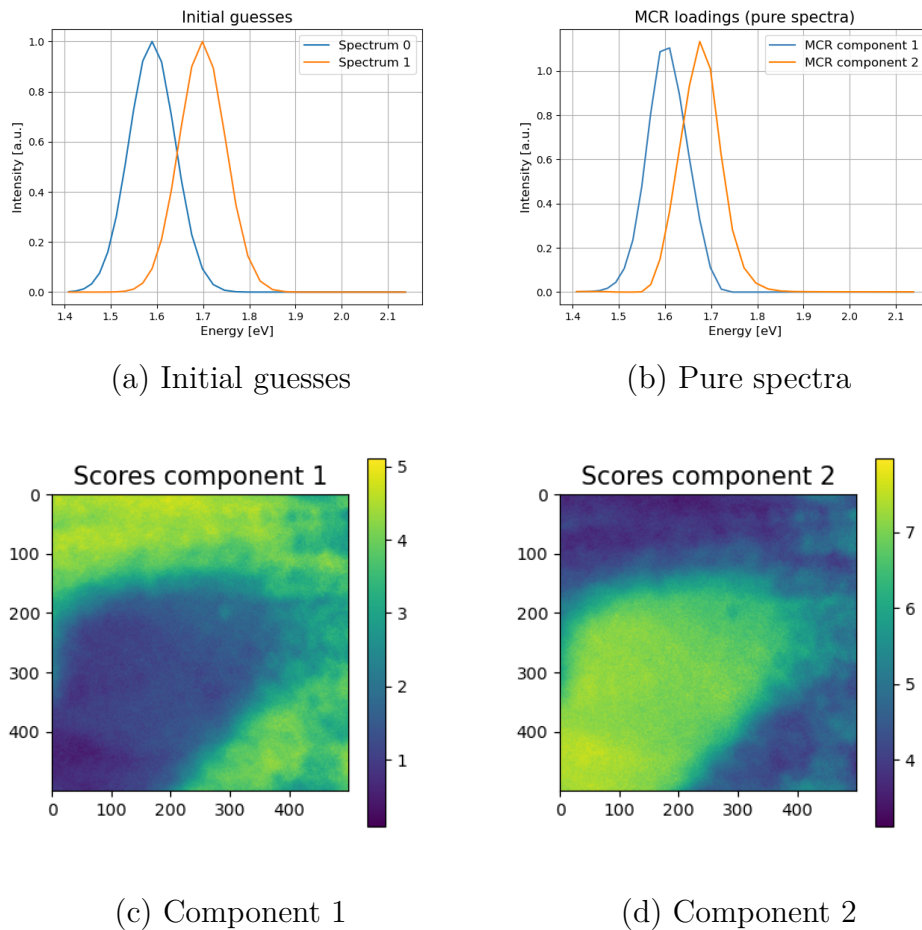
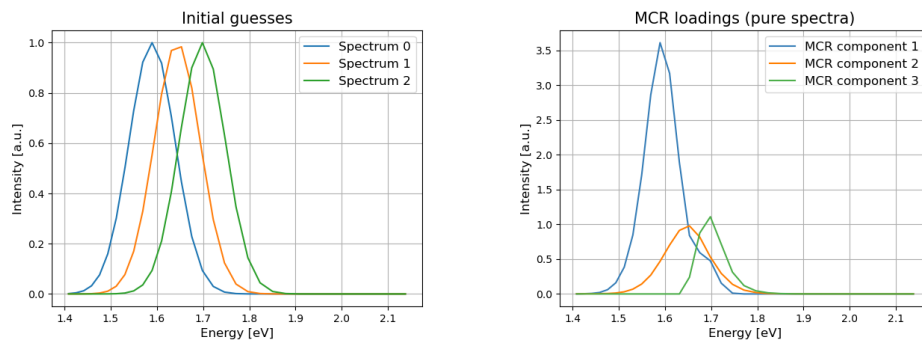


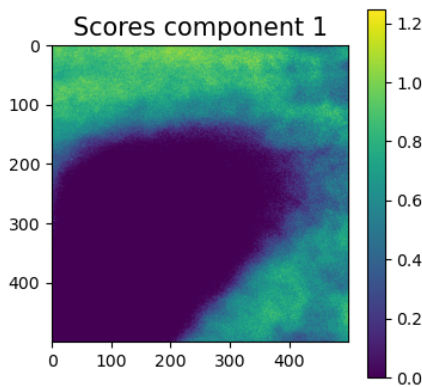
Figure C.5: Results from MCR with two components - the initial guesses of the spectra in (a), the pure spectra returned by the MCR algorithm in (b), and the scores/contractions of each component in (c) and (d).

C.3.2 Using 3 MCR components (without regularization)

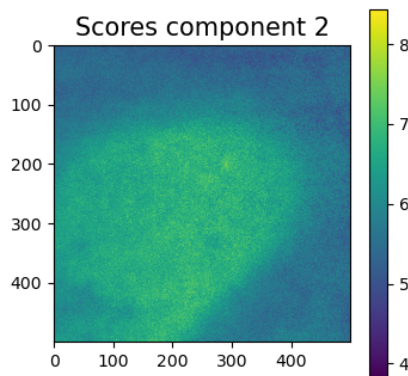


(a) Initial guesses

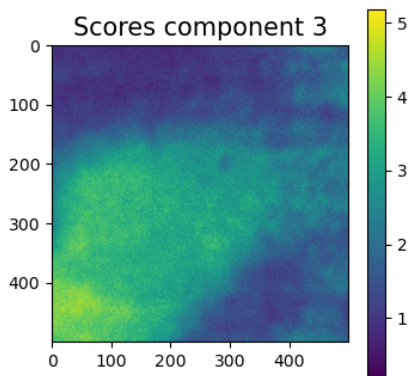
(b) Pure spectra



(c) Component 1



(d) Component 2



(e) Component 3

Figure C.6: Results from MCR with three components - the initial guesses of the spectra in (a), the pure spectra returned by the MCR algorithm in (b), and the scores/contractions of each component in (c), (d) and (e).

**C.3.3 Using 4 MCR components (without regularization)**

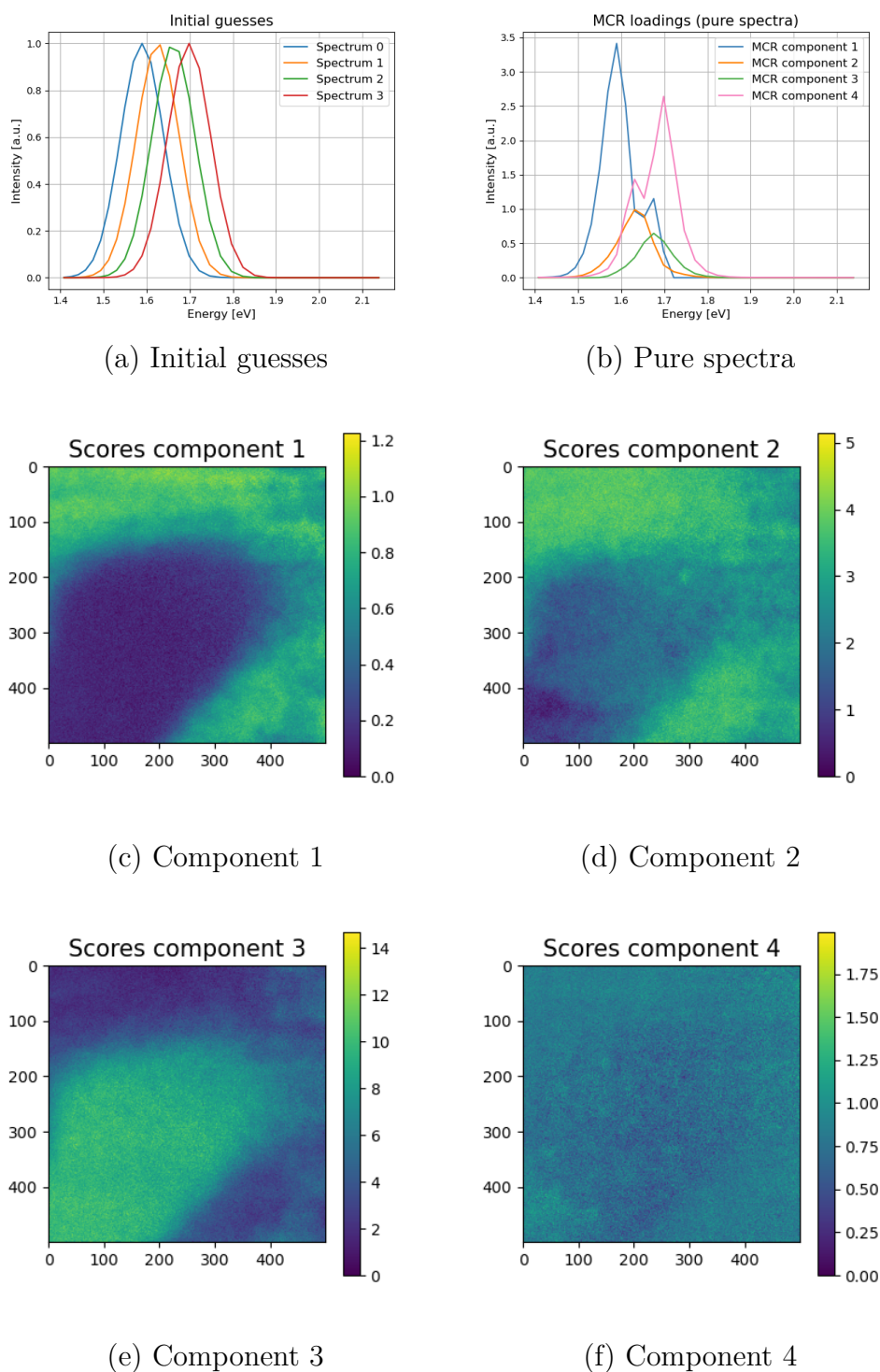
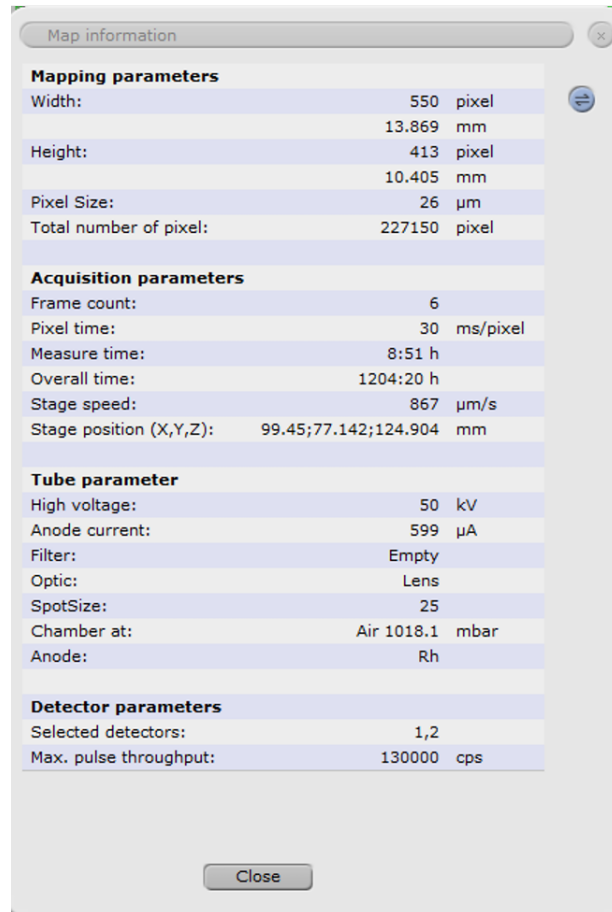


Figure C.7: Results from MCR with four components - the initial guesses of the spectra in (a), the pure spectra returned by the MCR algorithm in (b), and the scores/concentrations of each component in (c), (d), (e) and (f).

---

## D XRF supplementaries



The image shows a software window titled "Map information" with a close button (X) in the top right corner. The window contains four sections of parameters, each with a blue header and a list of key-value pairs. A "Close" button is located at the bottom center of the window.

Mapping parameters		
Width:	550	pixel
	13.869	mm
Height:	413	pixel
	10.405	mm
Pixel Size:	26	$\mu\text{m}$
Total number of pixel:	227150	pixel
Acquisition parameters		
Frame count:	6	
Pixel time:	30	ms/pixel
Measure time:	8:51	h
Overall time:	1204:20	h
Stage speed:	867	$\mu\text{m/s}$
Stage position (X,Y,Z):	99.45;77.142;124.904	mm
Tube parameter		
High voltage:	50	kV
Anode current:	599	$\mu\text{A}$
Filter:	Empty	
Optic:	Lens	
SpotSize:	25	
Chamber at:	Air 1018.1	mbar
Anode:	Rh	
Detector parameters		
Selected detectors:	1,2	
Max. pulse throughput:	130000	cps

Figure D.1: *Settings used for obtaining the XRF images in this work.*



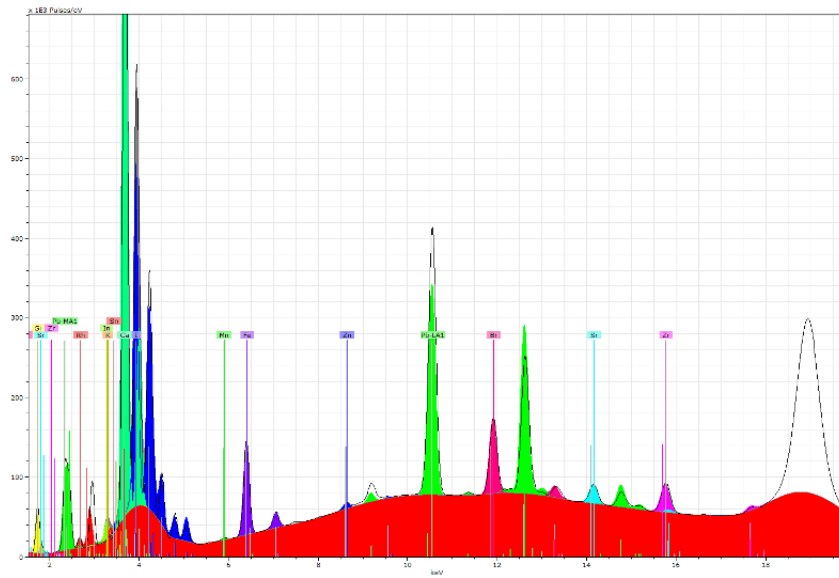


Figure D.2: Recorded XRF spectrum for sample 1.

Results					
Element	AN	Net	norm. C. [wt.%]	Atom C. [at.%]	
Rhodium	45	34779239	0.00	0.00	
Manganese	25	480614	0.04	0.03	
Zinc	30	1529890	0.05	0.03	
Strontium	38	6354633	0.21	0.11	
Zirconium	40	9595152	0.40	0.20	
Silicon	14	6086742	30.56	49.12	
Potassium	19	0	0.00	0.00	
Calcium	20	210191687	33.79	38.07	
Iron	26	20991113	1.18	0.96	
Bromine	35	22263137	0.66	0.37	
Indium	49	195756	0.23	0.09	
Tin	50	122464	0.20	0.08	
Lead	82	110224081	4.86	1.06	
Iodine	53	117851820	27.81	9.89	
<b>Total</b>			<b>100.00</b>	<b>100.00</b>	

Figure D.3: The XRF software's estimated concentrations for sample 1.

---

## E Supplementary material to Section 5.2

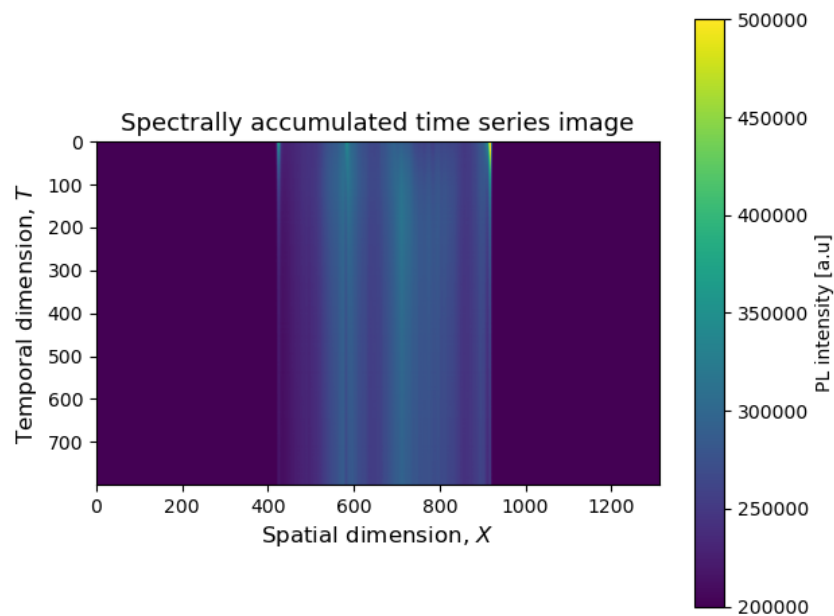


Figure E.1: *The spectrally accumulated time series image of sample 2. It is obtained by summing the counts in all wavelength bands for every pixel  $(X, T)$  in the hyperspectral image.*

---

## F Supplementary material to Section 5.7

### Examining pixel spectra in region D

For examining the information hidden in the individual pixels in region D, some selected pixel spectra are shown in Figure F.1. The spectra are generally very similar to the region's mean spectrum (see Figure 5.18). However, some of the pixel spectra appear noisy at around 1.56-1.57 eV, which explains why region D's mean spectrum in Figure 5.18 is less smooth in this spectral range. Apart from this, the mean spectrum seems representative of the entire region.

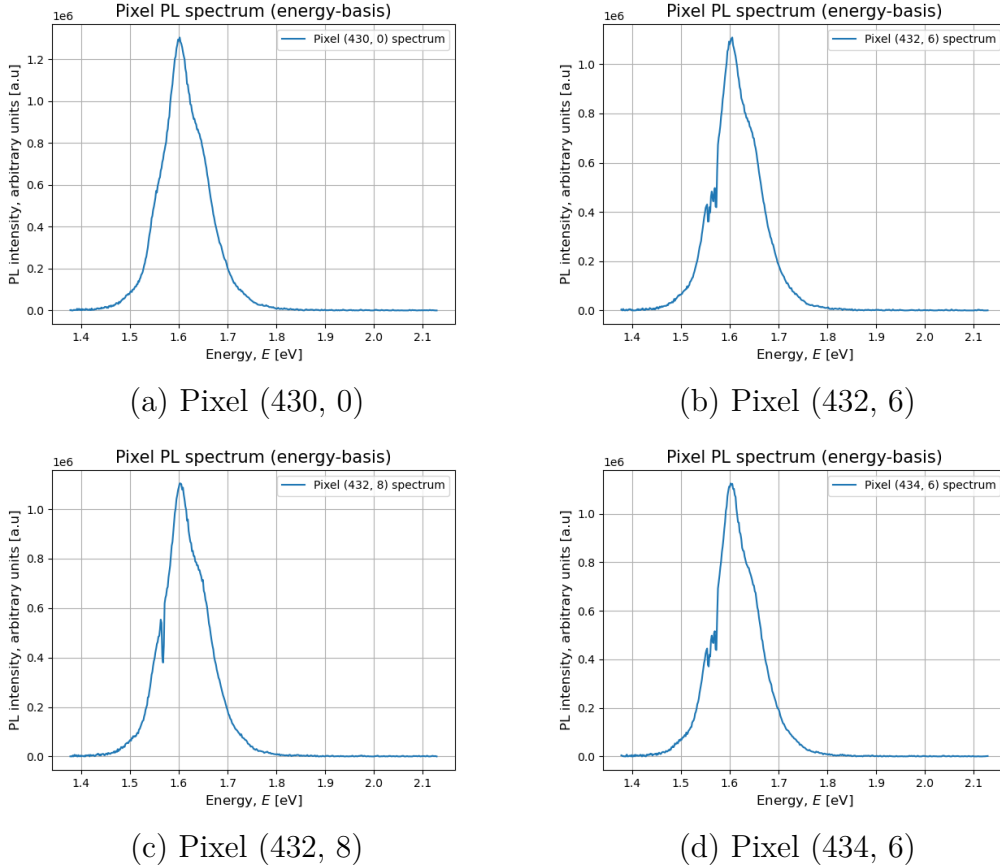


Figure F.1: Selected pixel spectra from region D in Figure 5.17. As can be seen in (b), (c), and (d), some of the pixel spectra in this region appear noisy at around 1.56-1.57 eV.

---

---

## G Supplementary material to Section 5.8

### G.1 Motivating regularized regression in MCR

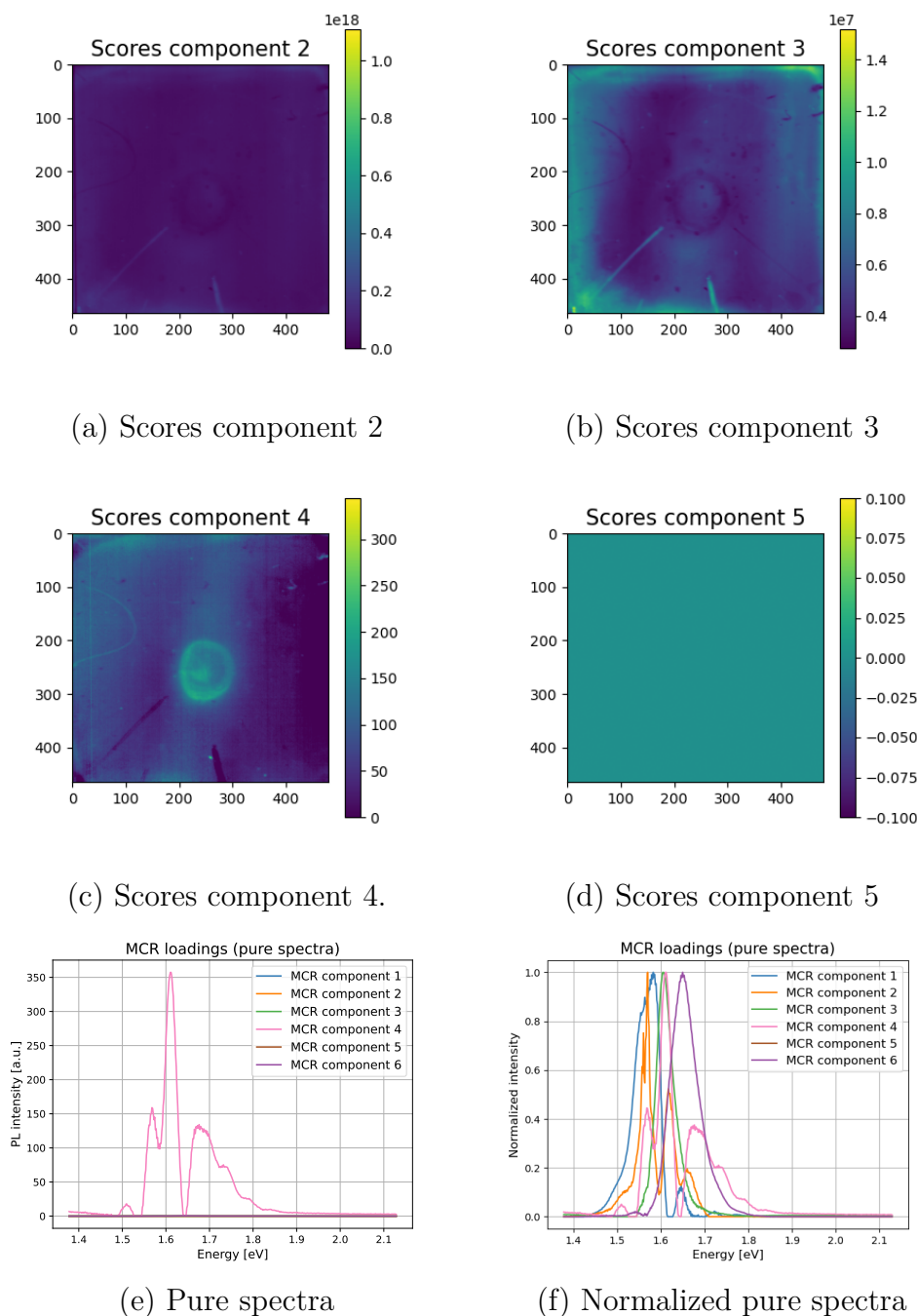
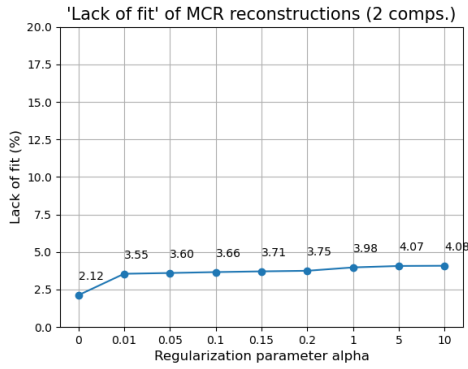
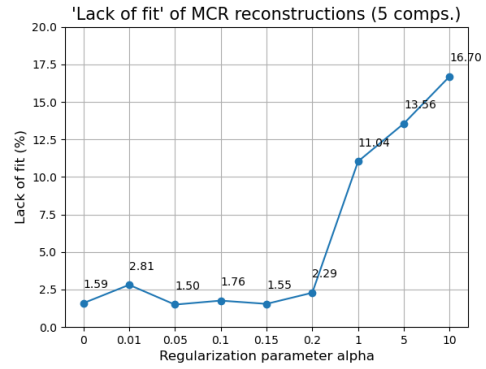


Figure G.1: Selected score images and the pure components of a non-regularized MCR model with six components. Figures (a)-(d) show score images (concentrations) of selected components. Figure (e) shows the loading plot where all pure spectra are plotted in the same coordinate system. Figure (f) shows the normalized pure spectra, where all spectra have been divided by their maximum value.

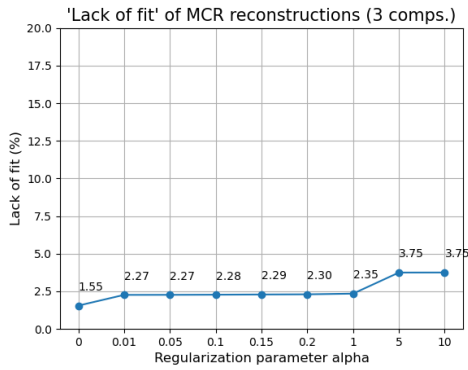
## G.2 "Lack of fit" as a function of regularization strength



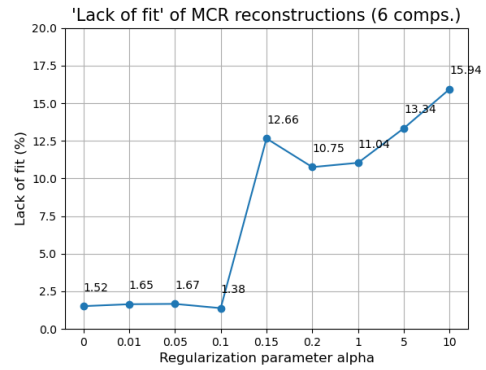
(a) 2 comps.



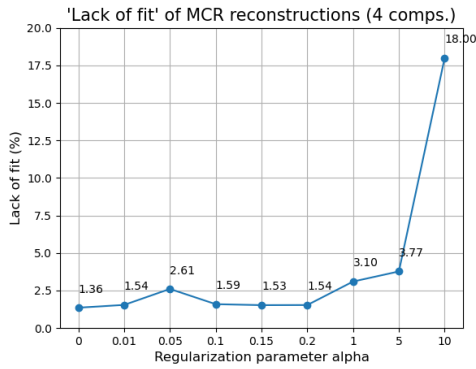
(d) 5 comps.



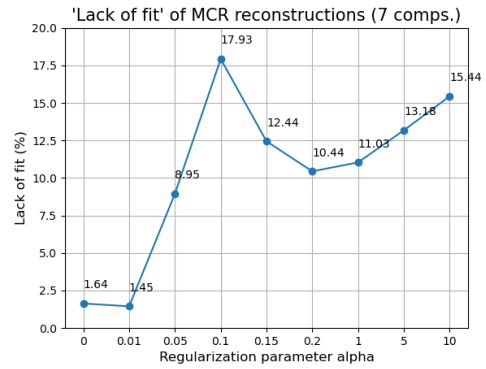
(b) 3 comps.



(e) 6 comps.



(c) 4 comps.



(f) 7 comps.

Figure G.2: "Lack of fit" (LoF) as a function of the regularization parameter  $\alpha$  for MCR models with different numbers of components.

### G.3 Pure spectra and concentrations for the selected MCR models with three and four components

## G.3 Pure spectra and concentrations for the selected MCR models with three and four components

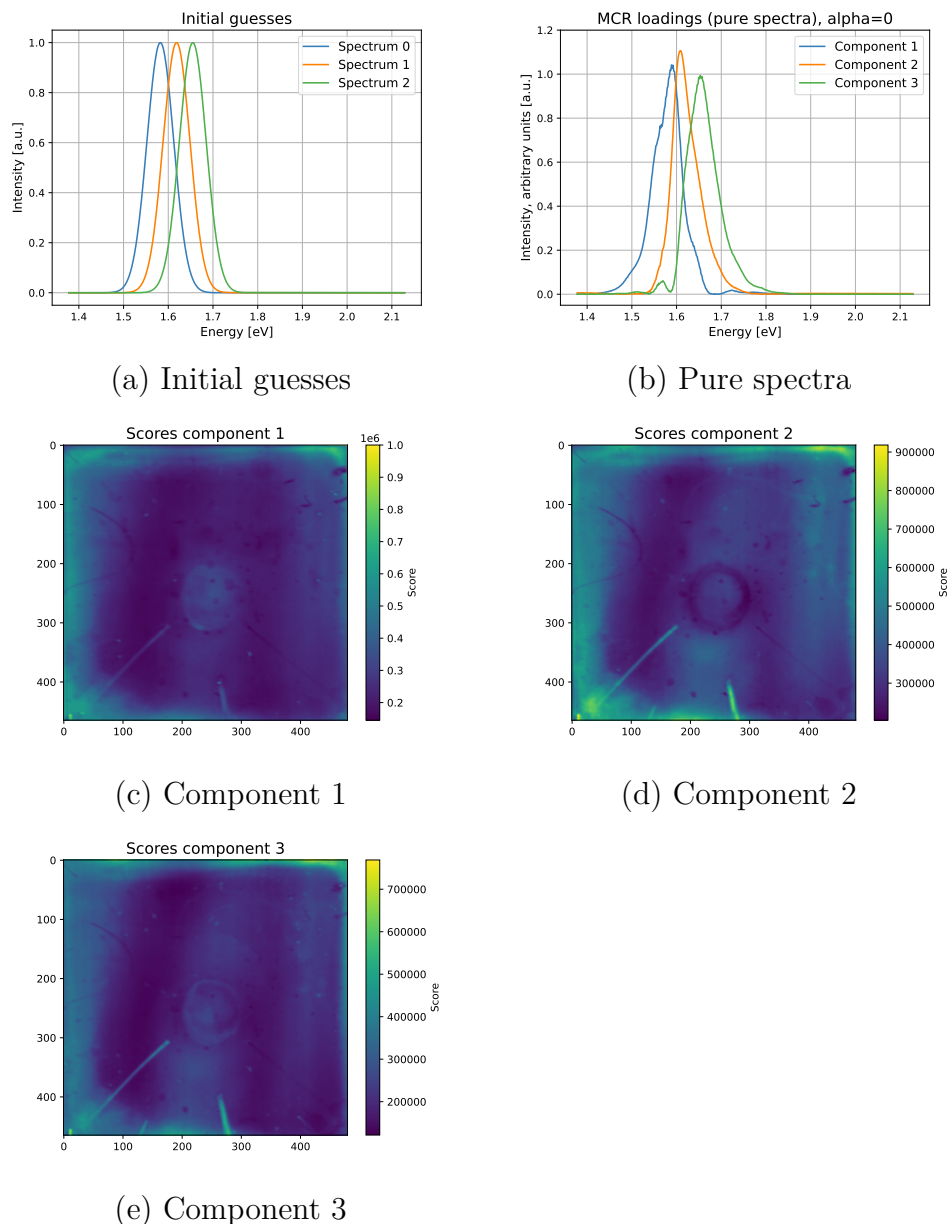


Figure G.3: Results from MCR with three components and  $\alpha = 0$  - the initial guesses of the spectra in (a), the pure spectra returned by the MCR algorithm in (b), and the scores/concentrations of each component in (c), (d) and (e).

### G.3 Pure spectra and concentrations for the selected MCR models with three and four components

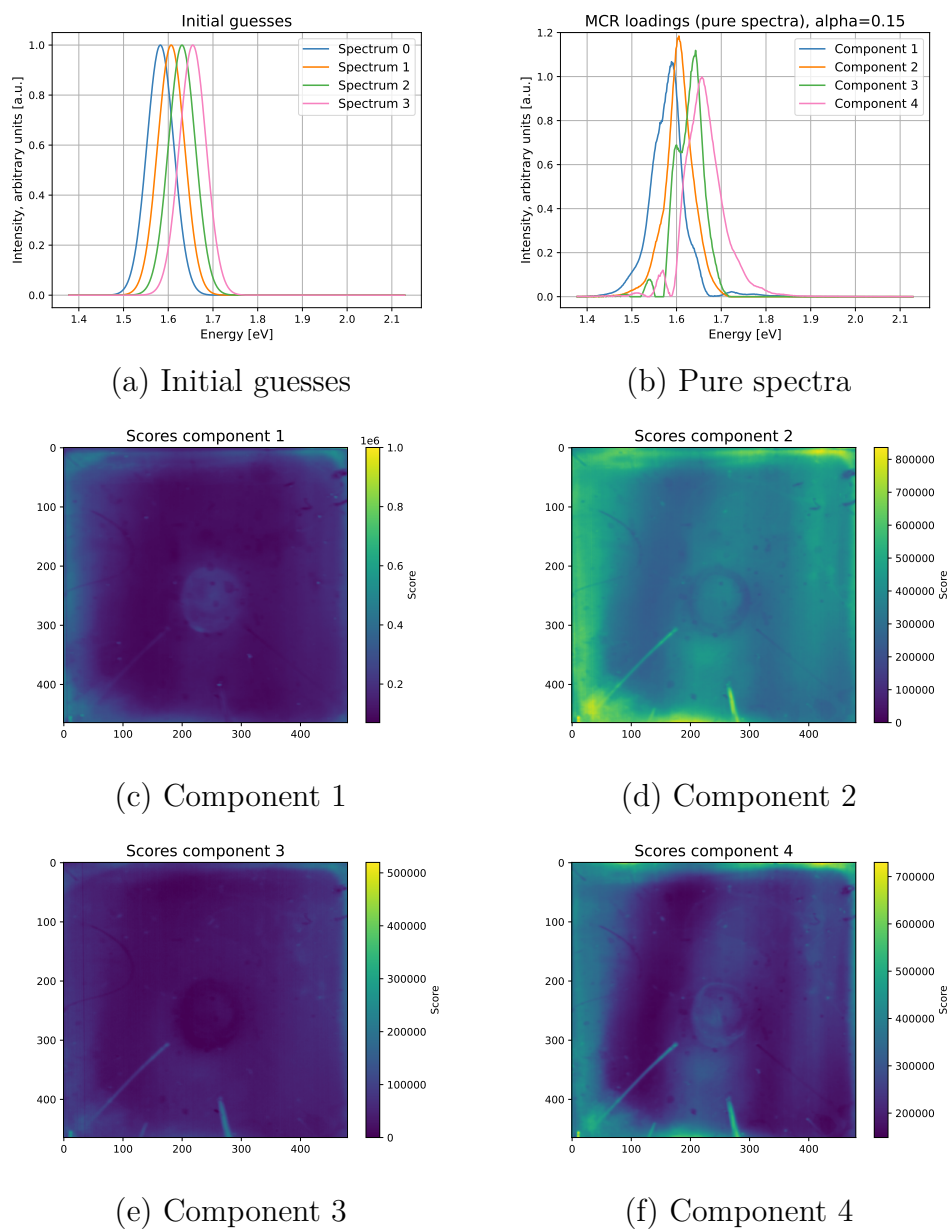


Figure G.4: Results from MCR with four components and  $\alpha = 0.15$  - the initial guesses of the spectra in (a), the pure spectra returned by the MCR algorithm in (b), and the scores/concentrations of each component in (c), (d), (e) and (f).



### G.4 Score images for MCR model with five components and $\alpha = 0.15$

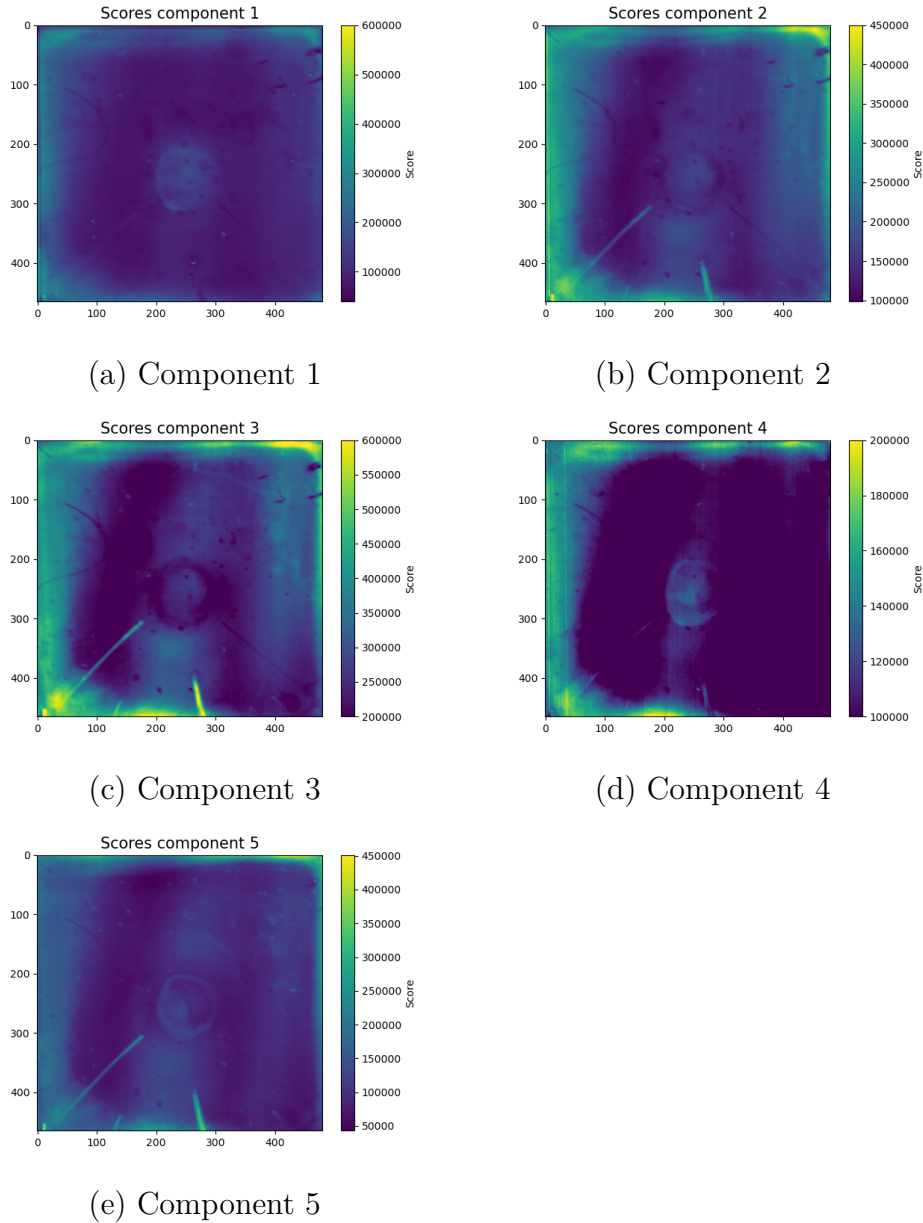


Figure G.5: *Score images for the MCR model with five components and  $\alpha = 0.15$ . Vertical stripes can be observed in (a), (b), (d), and (e), which may be attributed to noise from the camera.*

## G.5 Four-component MCR models with different regularization strengths

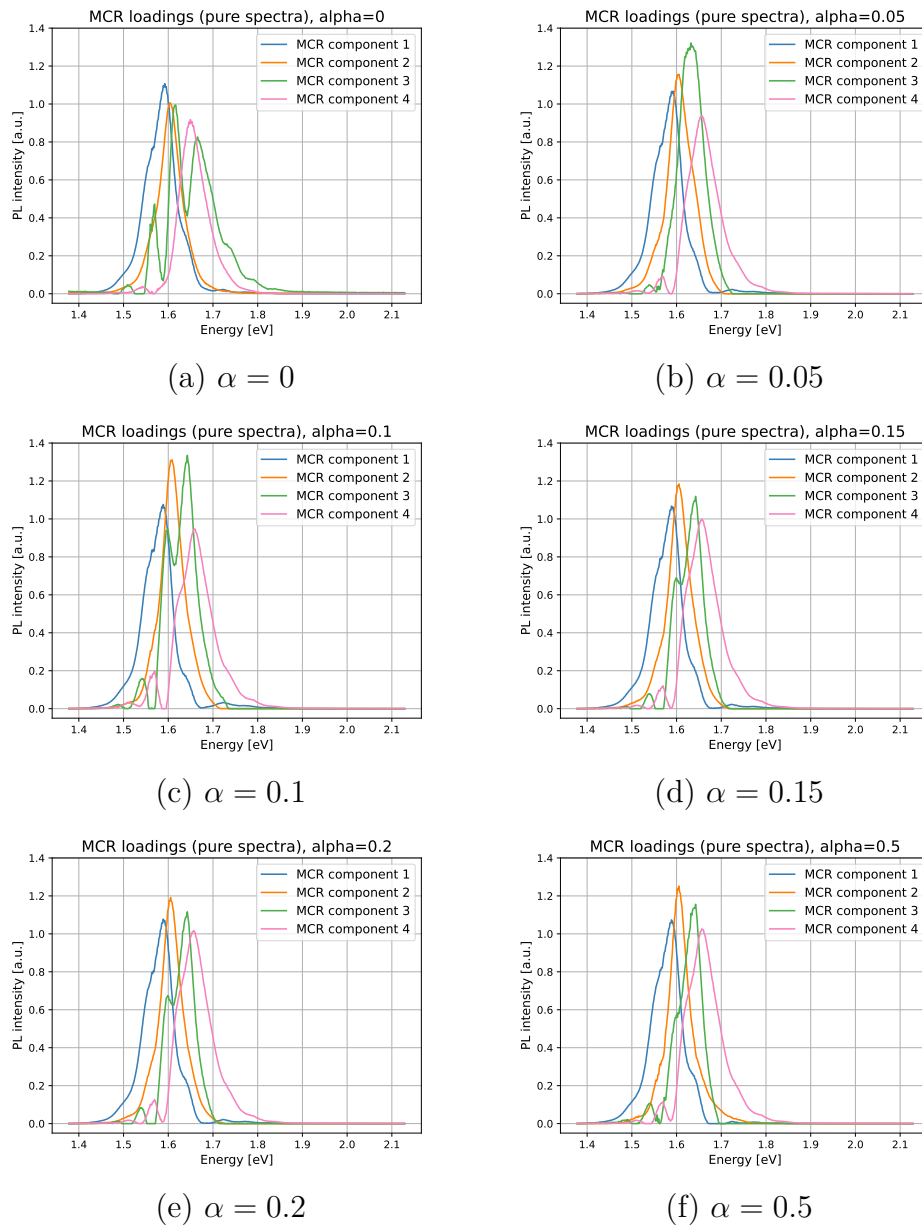
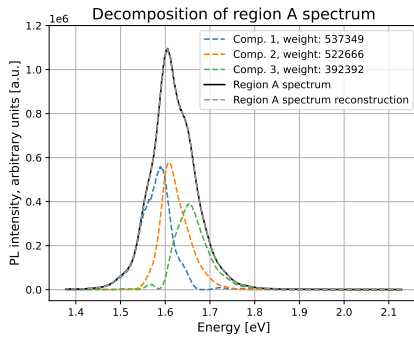


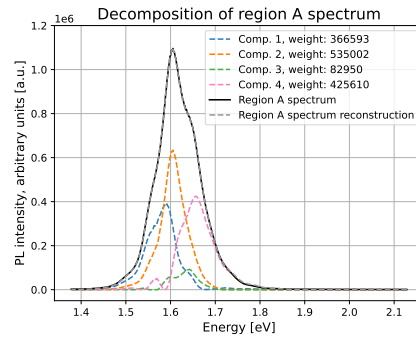
Figure G.6: *Pure spectra of MCR models with four components and different regularization strengths (different values of  $\alpha$ ). The locations of the spectra's PL peaks are almost identical for models with  $\alpha > 0$ .*

## G.6 Decompositions and reconstructions of mean PL spectra for regions A, B, C, and D

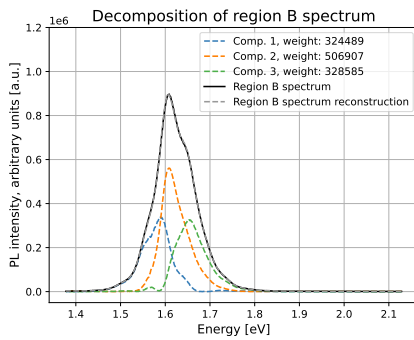
## G.6 Decompositions and reconstructions of mean PL spectra for regions A, B, C, and D



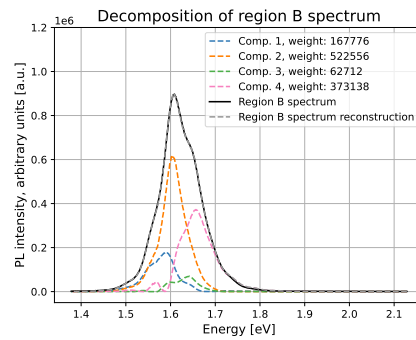
(a) Region A (3 comps.)



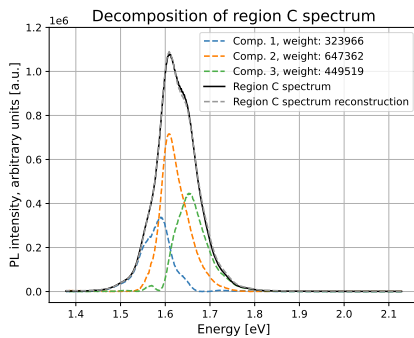
(b) Region A (4 comps.)



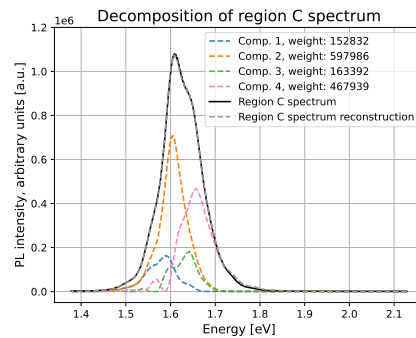
(c) Region B (3 comps.)



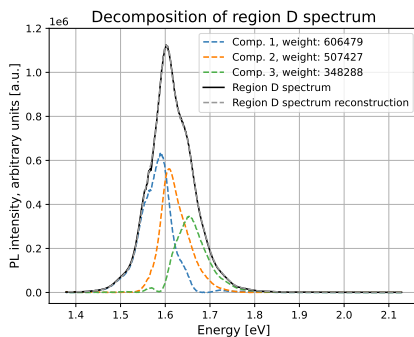
(d) Region B (4 comps.)



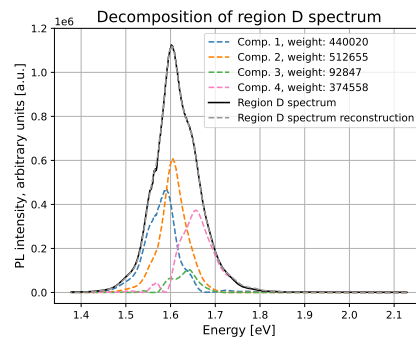
(e) Region C (3 comps.)



(f) Region C (4 comps.)



(g) Region D (3 comps.)



(h) Region D (4 comps.)

Figure G.7: Decompositions and reconstructions of the mean PL spectra for regions A, B, C, and D, obtained using the three-component MCR model with  $\alpha = 0$  and the four-component model with  $\alpha = 0.15$ .





**Norges miljø- og biovitenskapelige universitet**  
Noregs miljø- og biovitenskapelige universitet  
Norwegian University of Life Sciences

Postboks 5003  
NO-1432 Ås  
Norway