



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA STROJNÍHO INŽENÝRSTVÍ

FACULTY OF MECHANICAL ENGINEERING

ÚSTAV MATEMATIKY

INSTITUTE OF MATHEMATICS

**VIZUALIZACE SPEKTROSKOPICKÝCH DAT POMOCÍ
METODY ANALÝZY HLAVNÍCH KOMPONENT**

VISUALIZATION OF SPECTROSCOPIC DATA USING PRINCIPAL COMPONENT ANALYSIS

DIPLOMOVÁ PRÁCE

MASTER'S THESIS

AUTOR PRÁCE

AUTHOR

Bc. David Šrenk

VEDOUCÍ PRÁCE

SUPERVISOR

Ing. Pavel Pořízka, Ph.D.

BRNO 2018

Zadání diplomové práce

Ústav:	Ústav matematiky
Student:	Bc. David Šrenk
Studijní program:	Aplikované vědy v inženýrství
Studijní obor:	Matematické inženýrství
Vedoucí práce:	Ing. Pavel Pořízka, Ph.D.
Akademický rok:	2018/19

Ředitel ústavu Vám v souladu se zákonem č.111/1998 o vysokých školách a se Studijním a zkušebním řádem VUT v Brně určuje následující téma diplomové práce:

Vizualizace spektroskopických dat pomocí metody analýzy hlavních komponent

Stručná charakteristika problematiky úkolu:

Spektroskopie laserem buzeného plazmatu je vhodná pro určování prvkového složení neznámých vzorků nacházejících se v libovolném skupenství. Základním výstupem měření je emisní spektrum, které však bez následné úpravy a dodatečných informací neposkytuje intuitivní reprezentaci složení vzorku. Pomocí různých chemometrických metod, především analýza hlavních komponent – principal component analysis – PCA) lze vzorky na základě naměřených spekter vhodně vizualizovat a následně i dělit do tříd, které sdružují vzorky s podobnými vlastnostmi. Při analýze heterogenního materiálu dochází k ablaci na rozhraní dvou matic a tím i k propojení klastrů, což následně třídění znemožňuje a určité množství dat je nutné z analýzy odstranit. Výstupem této práce bude metodika pro určení bodů, které není možné jednoznačně přiřadit k jednotlivým třídám.

Cíle diplomové práce:

S (Principal Component Analysis, PCA) – a to především z pohledu vizualizace a snížení dimenze úkolu. Následně provede srovnání způsobů přípravy dat a jejich vliv na rozdělení bodů v PCA prostoru. Vhodným návrhem experimentu bude připraven datový soubor simulující ablaci na rozhraní. Tento datový soubor bude následně zpracováván nejen z pohledu matematicko–statistického, ale i z pohledu fyzikálně–chemického; tedy výsledky budou vyhodnoceny s ohledem na původ bodů v PCA prostoru a jejich relativní příslušnost k dané matici.

Seznam doporučené literatury:

MIZIOLEK, A. W., PALLESECHI, V. a SCHECHTER I. Laser Induced Breakdown Spectroscopy. 1. vydání. Cambridge: Cambridge University Press, 2006. ISBN 978-051-1541-261.

MELOUN, M. a MILITKÝ, J. Statistická analýza experimentálních dat. 2. vydání. Praha: Academia, 2004. ISBN 80-200-1254-0.

Termín odevzdání diplomové práce je stanoven časovým plánem akademického roku 2018/19

V Brně, dne

L. S.

prof. RNDr. Josef Šlapal, CSc.
ředitel ústavu

doc. Ing. Jaroslav Katolický, Ph.D.
děkan fakulty

Abstrakt

Diplomová práce se věnuje spektroskopii laserem buzeného plazmatu k určování prvkového složení neznámých vzorků. Bylo nutné navrhnout vhodnou metodu pro kvalifikaci materiálu pomocí emisních spekter indukovaných laserem. Pro kvalifikaci prvkového složení bylo nezbytné použít různé chemometrické metody a předzpracování dat. Kýžené kvalifikace jsme dosáhli projekcí dat do nového PCA prostoru, následného vytvoření shluků a vymezení Euklidovských vzdáleností jednotlivých shluků mezi sebou.

Experiment v praktické části byl nastaven na detekci rozhraní dvou prvků. Vytvořili jsme tedy datový soubor simulující ablaci na rozhraní. Tento datový soubor byl postupně zpracován matematicko-chemicko-fyzikálním pohledem. Bylo zpracováno několik postupů pro analýzu dat: aproximace pomocí Lorenzovy, Gaussovy a Voigtovy funkce a také metody předzpracování, jako detekce odlehlých bodů, standardizace dat různými způsoby a následné použití metody hlavních komponent. V práci je zaznamenán soubor shrnutí postupů pro kvalifikaci vstupních dat.

Abstract

This diploma thesis deals with using laser-induced breakdown plasma spectroscopy for determining the elemental structure of unknown samples. It was necessary to design an appropriate method to qualify material by laser-induced emission spectrum. Pretreatment of data and using a variety of chemometrics methods had to be done in order to qualify the structure of elements. We achieved a required solution by projecting the data to a new PCA space, creating clusters and computing the Euclidean distance between each cluster.

The experiment in the practical part was set to detect an interface of two elements. We created a data file simulating the ablation on the interface. This data set was gradually processed applying a mathematical-chemical-physical view. Several data procedures have been compiled: approximation by Lorenz, Gauss and Voigt function and also a pretreatment method such as the detection of outliers, standardization by several procedures and subsequent use of principal components analysis. A summarization of processes for input data is fully described in the thesis.

klíčová slova

metoda hlavních komponent (PCA), spektroskopie laserem buzeného plazmatu (LIBS), vizualizace

keywords

principal component analysis (PCA), laser-induced breakdown spectroscopy (LIBS), visualization

Šrenk, D.: *Vizualizace spektroskopických dat pomocí metody analýzy hlavních komponent*, Brno: Vysoké učení technické v Brně, Fakulta strojního inženýrství, 2019. 63 s. Vedoucí diplomové práce Ing. Pavel Pořízka Ph.D.

Prohlašuji, že jsem diplomovou práci *Vizualizace spektroskopických dat pomocí metody analýzy hlavních komponent* vypracoval samostatně pod vedením Ing. Pavla Pořízky, Ph.D. s použitím materiálů uvedených v seznamu literatury.

David Šrenk

Tímto bych rád poděkoval především vedoucímu mé diplomové práce Ing. Pavlu Pořízkovi, Ph. D. za konzultace, odborné vedení a cenné rady nutné k vypracování této práce.

David Šrenk

Obsah

Úvod	13
1 LIBS	14
1.1 Sestava	14
1.2 Laserem buzené plazma	15
1.3 Intenzita Spektrální čáry	17
1.3.1 Z pohledu kvantové fyziky	17
1.3.2 Interpretace spekter	19
1.4 Rozšíření spektrálních čar	20
1.4.1 Dopplerovské rozšíření	21
1.4.2 Starkovo rozšíření (tlakové rozšíření)	22
1.4.3 Voigtův profil	23
2 Předzpracování dat	25
2.1 Detekce outlierů	25
2.1.1 Detekce založená na blízkosti	25
2.2 Redukce dimenze	27
2.3 Scaling	27
2.3.1 Centrování zdrojové matice X	27
2.3.2 Standardizace směrodatnou odchylkou	27
2.3.3 MinMax standardizace	28
2.4 Odečtení pozadí	28
3 PCA	29
3.1 Obecně	29
3.2 Matematický popis	30
3.3 Získání hlavních komponent	30
3.3.1 Alternativní zavedení optimalizačního problému	32
3.4 Predikce na základě modelu	34
3.5 Grafická reprezentace výstupu PCA	35
3.5.1 Získání počtu hlavních komponent	35
3.5.2 Graf komponentních vah (Plot Components Weights)	36
3.5.3 Rozptylový diagram komponentního skóre (Scatterplot)	36
3.5.4 Dvojný graf(Biplot)	38
4 Praktická část	39
4.1 Popis a detekce vzorku	40
4.2 Přístroj	40
4.3 Data	43
4.4 Experiment	44
4.4.1 Test normality	44
4.4.2 PCA	45
4.4.3 Korelační analýza	52
4.4.4 Alternativní přístup	53
5 Závěr	58

Úvod

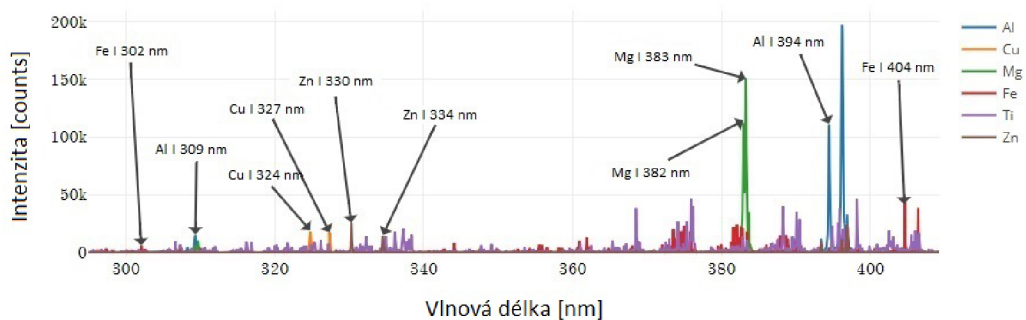
Práce se skládá z teoretické a praktické části. V teoretické části v kapitole LIBS práce pojednává o spektroskopii laserem buzeného plazmatu, jedná se o metodu měření emisní spektroskopie. Principem této metody je zaostřit laserový paprsek na povrch vzorku, kde dochází k přenosu energie a tvorbě plazmatu. Toto záření je optickou soustavou přenášeno do spektrometru a následně analyzováno detektorem. LIBS tedy poskytuje prvkovou informaci o složení vzorku, kvalitativní i kvantitativní. Tato metoda je založena na rozdílné spektrální charakteristice všech vzorků. Následně pojednává o fyzikálních vlastnostech plazmatu jako je například elektronová teplota a také o Dopplerovském či Starkově rozšíření spektrálních čar. Následuje teoretická kapitola předzpracování dat kde se pojednává o detekci outlierů pomocí detekce založené na blízkou, redukce dimenze, odečtení pozadí, scaling neboli standardizace pomocí směrodatné odchylky či MinMax standardizace. Poslední nejrozsáhlejší teoretická kapitola PCA, neboli metoda analýzy hlavních komponent vhodná pro zpracování vícerozměrných dat. Cílem metody je získat vzájemnou korelaci původních dat a vytvořit nový méně dimenzionální PCA prostor, kde můžeme vhodně interpretovat výsledky. Je zde zaveden matematický popis pro získání hlavních komponent, které vystupují jako nové proměnné v PCA prostoru, jedná se o optimalizační problém maximalizace rozptylu. Tento problém můžeme alternativně zavést jako minimalizační úlohu projekce ve smyslu nejmenších reziduálních čtverců. Zavedeme prediktivní model a následné grafické reprezentace výstupu metody hlavních komponent. V praktické části využijeme teoretických poznatků a pomocí metody LIBS zkompletujeme vstupní data pro následnou analýzu, při které aplikujeme algoritmy na předzpracování dat, detekci a vizualizaci pomocí metody hlavních komponent. Experiment prováděný v praktické části se snaží detekovat a vizualizovat rozhraní dvou prvků. Hledáme oblasti v novém PCA prostoru, kterým přiřadíme prvkové složení a také hranice mezi těmito oblastmi.

1 LIBS

Laser induced breakdown spectroscopy (=spektroskopie laserem buzeného plazmatu). V článku od Dr. Jamese Winefordnera z roku 2012 byla LIBS popsána jako "a future super star" [9] ve srovnání s jinými analytickými metodami. Jedná se o jedinou technologii, která je použitelná na detekci tak široké škály vzorků. Lze analyzovat látky jakéhokoliv skupenství. Vzorek může být pevná, kapalná i plynná látka.

Před měřením je vhodné si vzorek předpřipravit, obecně však lze měřit vzorek bez přípravy. Není nutný žádný fyzický kontakt se vzorkem. Matici můžeme analyzovat ve vzduchu, tedy za atmosférického tlaku. Vzorek jsme schopni detekovat na různé vzdálenosti. Lze vyrobit kompaktní přenosný LIBS systém, ale také systém na dálkovou detekci. Metoda LIBS má poměrně dobrý výkon pro analýzu na blízko, jejíž výsledky jsou srovnatelné, ba dokonce i o řád lepší než rentgenové metody[27]. Reakční doba metody je kratší než 1 sekunda. V neposlední řadě je výhodou jednoduchá realizace měřicí aparatury a také levný provoz. Používá se v široké škále odvětví od medicíny přes průmysl až po archeologii. Jednoduše řečeno, laserový paprsek je zaostřen do úzkého bodu na povrchu vzorku a dojde k odpaření a vytvoření plazmatu. Jako výstup analýzy je detekce charakteristického záření laserem buzeného plazmatu, jehož spektroskopická analýza vede na čárové emisní spektrum. LIBS tedy poskytuje prvkovou informaci o složení vzorku získanou z optické emise, tedy kvalitativní a kvantitativní informaci. Kvalitativní analýza identifikuje jednotlivé prvky a kvantitativní analýza stanovuje obsah jednotlivých prvků. Umístění emisních čar určuje prvek a intenzita čar umožňuje kvantifikaci. Složení cílového materiálu je získáno z LIBS spektra za pomoci kalibračních křivek, neboli databáze emisních čar jednotlivých spekter. Tato metoda je založena na rozdílné spektrální charakteristice všech chemických prvků tzv. chemical fingerprint. Je tedy vhodná k určení prvkového složení vzorku.

Můžeme detekovat jak materiál skládající se z jednoho prvku, tak také směs více prvků. Výsledná spektrální čára je charakterizována stopami prvků obsažených v materiálu plus šumem, který vznikl při detekci signálu. Na obrázku můžeme vidět významné spektrální čáry daných prvků. Tato kapitola čerpá poznatky primárně z literatury [8],[20].

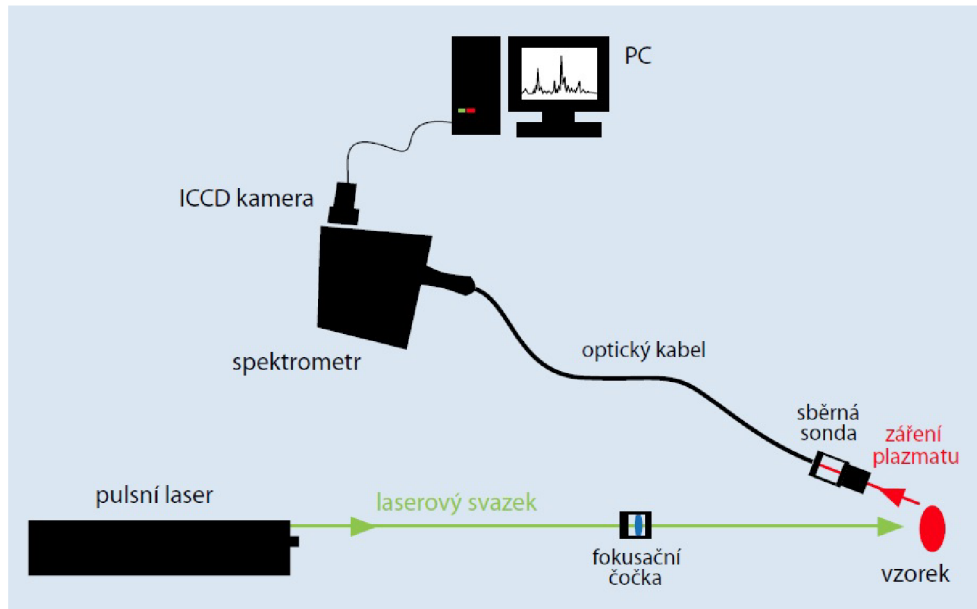


Obrázek 1: Spektra šesti prvků - Al, Cu, Mg, Zn, Fe, Ti

1.1 Sestava

První práce zabývající se analýzou povrchu pomocí laserů pocházejí již z 60 let. Sestava LIBS je jednoduchá a flexibilní, díky čemuž je možné sestavit přenosný LIBS systém, nebo LIBS s dálkovou detekcí. Detekční limity se pohybují pro většinu prvků v jednotkách $\frac{mg}{kg}$

a vzniklé krátery po ablací laserovým paprskem mají rozměry v desítkách μm až mm . Z tohoto pohledu vidíme, že metoda je téměř nedestruktivní a má díky tomu široké spektrum využití. Dílčí součásti, které potřebujeme pro sestavení jsou: pulzní laser, zaostřovací a sběrné optiky, spektrometr obsahující detektor, generátor pro časování pulzů a počítač pro zobrazení a vyhodnocení měření. Uspořádání jednotlivých součástí je znázorněno na obrázku 2 níže.



Obrázek 2: LIBS sestava, zdroj:[23]

Charakteristické vlastnosti laseru jsou: monochromaticnost (jedna vlnová délka v každém laserovém impulzu), polarizace (vlnění pouze v jedné rovině), koherence (světlo kmitá ve stejném směru, stejné frekvenci a stejné fázi), nondivergence (malá rozbíhavost paprsků). Používají se různé typy laserů, lišící se různou dobou trvání pulzu či vlnovou délkou. V současné době nejpoužívanější laser na vybuzení plazmatu je Nd:YAG laser. Jedná se o pevnolátkový laser. Aktivním materiálem je izotropní krystal Yttrium-Aluminium Granátu ($\text{Y}_3\text{Al}_5\text{O}_{12}$) dopovaný ionty neodymu (Nd^{3+}). Buzen je nejčastěji diodovou nebo xenonovou výbojkou. Tento typ laseru ve svém základním uspořádání emituje záření o vlnové délce 1064nm .

U LIBS si musíme dát pozor na dodržení správných podmínek k získání rovnovážného plazmatu, nebo také saturace signálu u prvků s vysokou koncentrací. To se může ukázat jako omezení a dokonce nemusí umožnit kvantifikaci analytů daného vzorku.

Lokální rovnováha plazmatu (LTE) předpokládá, že v malém okolí každého bodu je splněna podmínka termodynamické rovnováhy. Gradient teploty je nulový a intenzita ve všech směrech stejná. Je splněna pokud je hustota tak velká, že srážkové procesy převažují nad radiačními. Pro LTE lze tedy vzít předpoklad rovnovážného rozložení energie.

1.2 Laserem buzené plazma

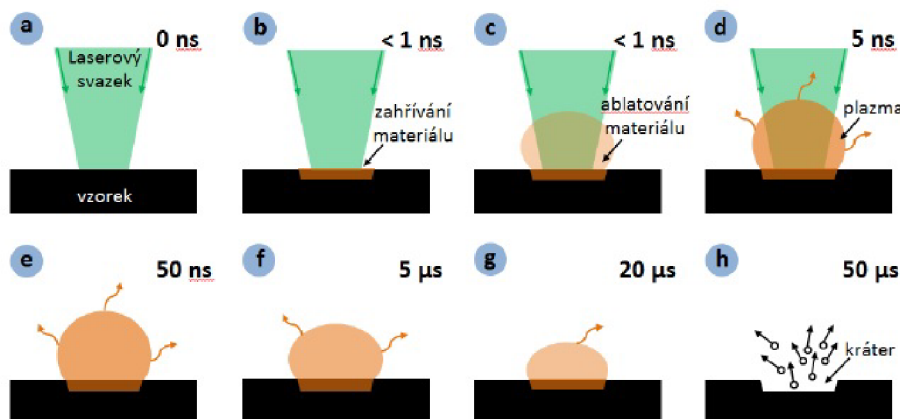
Jako plazma nazýváme silně ionizovaný plyn složený z iontů a elektronů, neutrálních atomů nebo molekul[28].

Ke vzniku plazmatu je zapotřebí přísun energie pomocí krátkého laserového pulsu zaostřeného optickou soustavou do ohniska. Podmínkou ke vzniku plazmatu je nutné překročit

určitou hodnotu plošné hustoty výkonu (irradiance), tedy dodání jednotkového množství energie na jednotku plochy za jednotku času. Tato hustota je pro každý materiál různá a nazývá se prahová hustota. Jedná se o irradianci, při které dojde ke vzniku plazmatu, které expanduje nadzvukovou rychlostí. S rostoucí energií laserového pulzu dochází k ablací většího množství materiálu. Výsledný signál je silnější, ale roste také pozadí signálu.

Při vzniku plazmatu jde o dva hlavní kroky vedoucí k porušení povrchu v důsledku optické excitace. Nejdříve se vytváří volné elektrony sloužící jako počáteční receptory energie. Dochází ke kolizi s neutrály a fotony. Druhým krokem je lavinová ionizace v ohnisku dopadu laserového paprsku. Volné elektrony jsou urychlovány elektrickým polem způsobeným optickým pulsem před kolizí. S rostoucí elektronovou energií dochází k termalizaci vlivem velkého počtu srážek, která produkuje lavinovou ionizaci, tedy vznik lavinové reakce. Ionizační energie jsou řádově $10eV$ [20].

Vývoj laserem buzeného plazmatu v čase a prostoru je znázorněn na obrázku 3. Laserový paprsek zaostřený optickou soustavou dopadá na povrch vzorku, povrch se začne zahřívat, vzniká ablace materiálu reagující s okolní atmosférou a vzniká plazma. To následně expanduje vně ve všech směrech od ohniska. V průběhu laserového záření roste hustota elektronů a plazmová elektronová frekvence do takové míry, až se samotné plazma stane pro laserový paprsek neprůhledné a laser začne předávat energii již ablatované svrchní vrstvě plazmatu. Toto indukované plazma se začne kuželovitě šířit ve směru zdroje paprsku. Zvuk s tímto procesem spojený je způsoben rázovou vlnou způsobenou rychlou expanzí z ohniska, tato rychlost je řádově v $\frac{km}{s}$. Měření studovaného materiálu provádíme se zpožděním, tzv. gate delay. Na začátku vývoje plazmatu nenese jeho záření cennou informaci o materiálu z důvodu brzdných a rekombinačních záření.

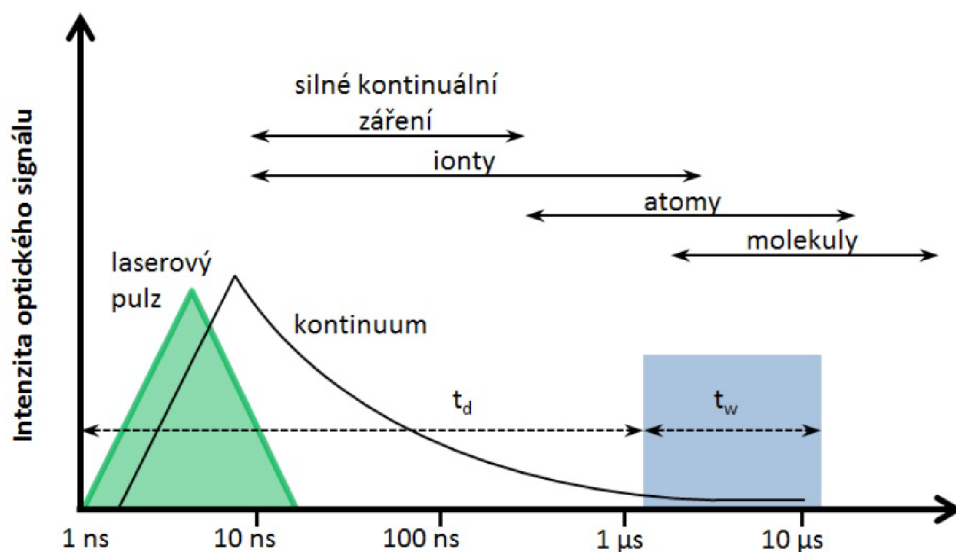


Obrázek 3: Cyklus vzniku plazmatu, zdroj: [26].

Při expanzi plazmatu dochází k ionizaci a excitaci částic na vyšší energetické hladiny. Tento jev trvá pouze krátkou dobu, poté plazma chladne a částice se vrací do svých základních stavů, při kterých emitují spektrální záření. Tuto emisi záření zaznamenáváme. Získáme čárové spektrum, které je pro každý prvek unikátní.

Na následujícím obrázku 4 vidíme emisi záření v závislosti na čase. Je zde znázorněn časový posun měření td (time delay) na hodnotu kolem $1\mu s$. Vidíme, že při této době by již mělo zanikat silné kontinuální záření a záření iontů obsažené v okolní atmosféře. Po době td by v emisním spektru měly být obsaženy čáry atomů, molekul a částečně iontů.

Tyto čáry jsou již poměrně dobře rozeznatelné. Délku časového okna měření nazveme integrační čas t_w (time width). Vhodný ke zvolení je interval $\langle 1\mu m, 10\mu m \rangle$.



Obrázek 4: časový režim měření, zdroj: [26]

1.3 Intenzita Spektrální čáry

Výstupem metody LIBS je charakterické spektrum nesoucí informaci o prvkovém složení vzorku. Jde o závislost intenzity záření emitovaného plazmatem na vlnové délce. Pomocí vlnové délky spektra získáme kvalitativní údaj, tedy jaké spektrální čáry daného prvku jsou ve vzorku zastoupeny. Po kalibraci přístroje, tedy kalibrace intenzity vůči množství sledovaného analytu, je možné odhadnout obsah jednotlivých prvků ve vzorku; jedná se tedy o kvantitativní analýzu. Každá spektrální čára je unikátním chemickým "otiskem prstů" daného materiálu. Spektrální čáry emitují od UV přes viditelnou část spektra až do IR oblasti. Je možné si vybrat libovolnou část spektra k analýze. Rozsah a jemnost spektrální čáry se odvíjí z vlastností daného měřicího přístroje vedoucí na objem dat výsledné matice.

1.3.1 Z pohledu kvantové fyziky

Atom je obklopen obíhajícími elektrony, které zaujímají různé diskrétní energetické hladiny. V plazmatu oproti jiným skupenstvím se částice poměrně rychle pohybují. Mají velkou kinetickou energii, díky které dochází ke vzájemné interakci částic. Pomocí de Broglieho teorie můžeme elektron popsat také jako vlnu:

$$\lambda = \frac{h}{p}, \quad (1.1)$$

kde h je Planckova konstanta p je hybnost částice a λ je jeho vlnová délka. Energie kvantových stavů elektronu je E_n je diskrétní veličina. Máme tedy diskrétní energetické stavy, v nichž se může elektron nacházet, známé také jako kvantové stavy. Pokud elektron přechází mezi energetickými hladinami, tak přijímá nebo uvolňuje určité diskrétní

Tabulka 1: Tabulka jednotlivých čar Al, Cu, Mg, Fe, Ti, Zn a jejich charakteristické spektrální čáry. Zdroj databáze NIST [38]

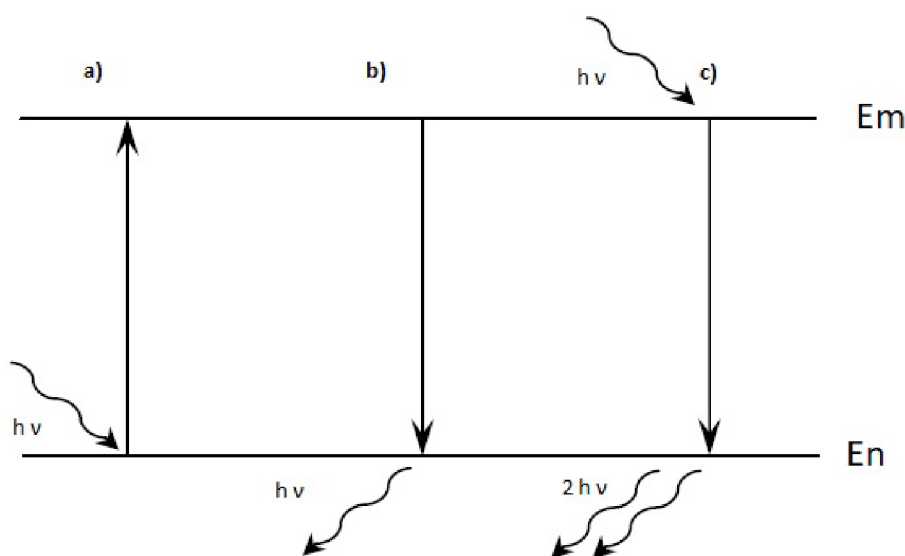
prvek	stupeň ionizace	vlnová délka[nm]	$A_{ki}[10^8 \cdot s^{-1}]$	$E_i[eV]$	$E_k[eV]$
Al	I	237,312	0,91	0,014	5,237
	I	308,215	0,59	0,000	4,022
	I	309,271	0,73	0,014	4,022
	I	394,401	0,50	0,000	3,143
Cu	I	324,754	1,40	0,000	3,817
	I	327,396	1,38	0,000	3,786
	I	510,554	0,02	1,389	3,817
	I	515,323	0,60	3,786	6,191
	I	521,820	0,75	3,817	6,191
Mg	II	279,553	2,60	0,000	4,434
	II	279,800	4,79	4,434	8,863
	II	280,270	2,57	0,000	4,422
	I	285,213	4,91	0,000	4,346
	I	382,935	0,90	2,709	5,946
	I	383,829	1,61	2,717	5,947
	I	518,360	0,56	2,717	5,108
Fe	I	253,720	3,70	2,404	7,289
	I	284,980	1,12	1,485	4,733
	I	302,060	0,76	0,000	4,103
	I	404,580	0,86	1,485	4,549
Ti	II	323,500	1,71	0,049	3,881
	II	334,900	1,61	0,607	4,308
	I	498,170	0,67	0,848	3,337
Zn	I	330,250	1,07	4,030	7,830
	I	334,500	1,50	4,078	7,784
	I	481,060	0,70	4,078	6,655

množství energie, které se nazývá kvantum. U LIBS je energie získávána ze vzájemné interakce částic. K excitaci dochází v důsledku tepelného pohybu částic. Existují přechody mezi energetickými hladinami atomů, které mohou probíhat třemi způsoby:

-Spontánní emise, jde o přechod z vyššího energetického stavu do nižšího, při které atom vyzáří foton o frekvenci ν splňující podmínku $E_2 - E_1 = h\nu$, kde h je Planckova konstanta.

-U absorpce elektron přijme foton o kvantu energie $h\nu$ a přejde z nižšího energetického stavu do vyššího.

-Stimulovaná emise je přechod elektronu z energeticky vyšší hladiny na energeticky nižší hladinu. Tento elektron byl stimulován fotonem o energii odpovídající tomuto přechodu $E_2 - E_1 = h\nu$, h je Planckova konstanta, ν je frekvence. Foton se nepohlí a pokračuje stejným směrem s emitovaným fotonem. Za vhodných podmínek můžeme spustit řetězovou reakci emisní stimulace.



Obrázek 5: Schéma přechdů mezi energetickými hladinami a) absorpce, b) spontánní emise, c) stimulované emise.

1.3.2 Interpretace spekter

Emisní spektrální čáry nesou také informaci o vlastnostech vzniklého plazmatu. Těmito vlastnostmi jsou mimo jiné teplota, kterou získáme z intenzity spektrální čáry a z rozšíření spektrální čáry získáme elektronovou hustotu. Kapitola vychází z [29] a [30].

Elektronová teplota

Elektrony, které přecházejí do excitovaných stavů a zpět do stavů základních, se pohybují po sférách kolem jádra. Mají tedy kinetickou energii, vedoucí na základě Boltzmannovi teorie k výpočtu elektronové teploty. Z této teploty jsme schopni vypočítat koncentraci elektronů. Intenzita přechodu spektrální čáry mezi dvěma hladinami m a n

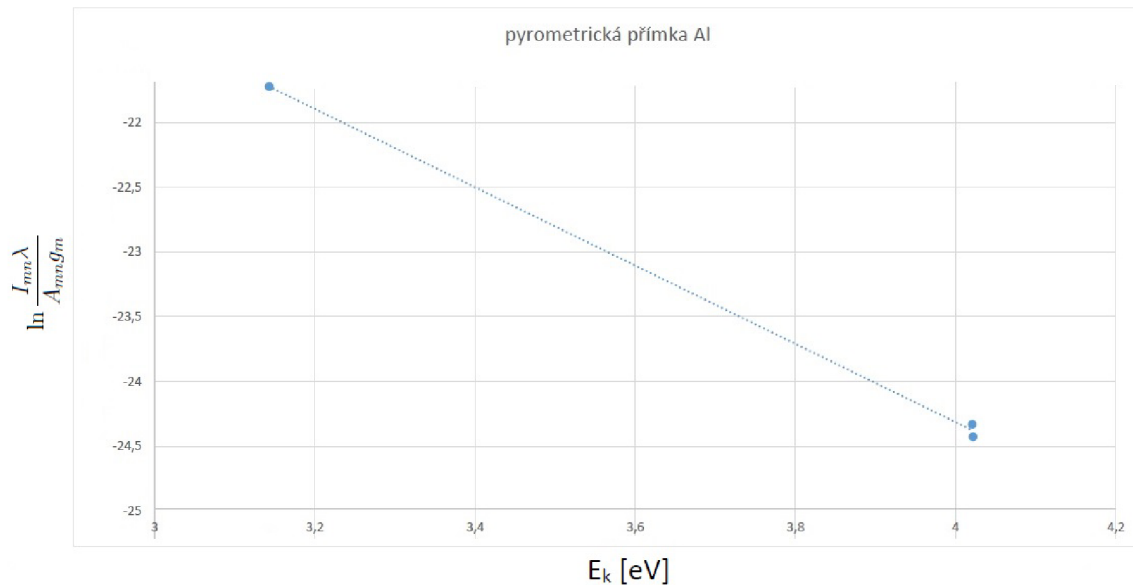
$$I_{mn} = A_{mn} N h \nu \frac{g_m}{\sigma} e^{-\frac{E_m}{kT_e}}. \quad (1.2)$$

A_{mn} je Einsteinův koeficient pravděpodobnosti spontánní emise, N je celkový počet excitovaných elektronů, ν je vlnová délka ($\frac{1}{\lambda}$), h je Planckova konstanta, g_m je stupeň degenerace,

σ je stavová suma, k je Boltzmannova konstanta, E_m je excitační energie hladiny m a T_e je teplota elektronů. Tento výraz můžeme zlogaritmovat a získáme rovnici pyrometrické přímky

$$\ln \frac{I_{mn}\lambda}{A_{mn}g_m} = -\frac{1}{kT_e}E_m + konst. \quad (1.3)$$

Vykreslíme-li tento logaritmus do grafu se závislostí na E_m , dostaneme přímku. Ze směrnice této přímky můžeme spočítat teplotu T_e . Pro lepší představu si postup ukážeme na hliníkovém vzorku, pomocí jeho tří významných čar. Přímku na obrázku jsme získali



Obrázek 6: Výpočet pyrometrické přímky pro hliník pomocí třech spektrálních čar.

pomocí lineární regrese. Rovnice této přímky je

$$y = -1,90054 \cdot 10^{19}x - 12,208. \quad (1.4)$$

Odtud elektronovou teplotu získáme pomocí vztahu:

$$T_e = -\frac{1}{(k \cdot (-1,90054 \cdot 10^{19}))} = 3830K. \quad (1.5)$$

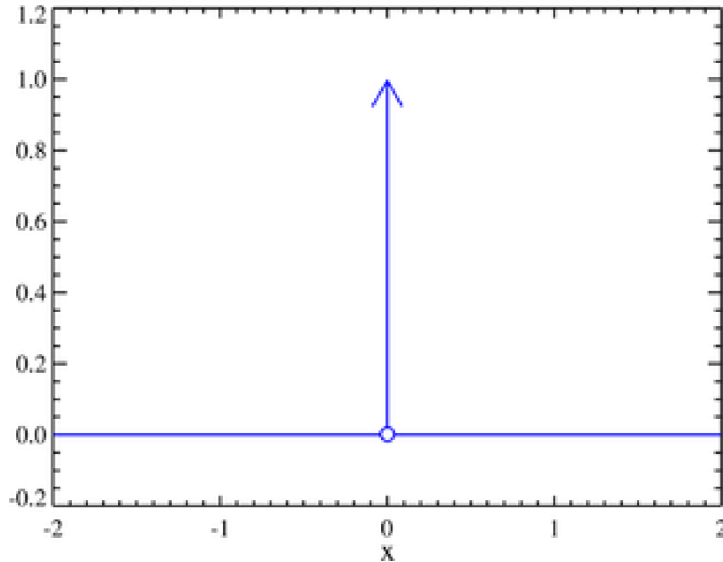
Měření teploty je pomocí této technologie náročné a slouží primárně k získání informace o teplotě v řádu jednotek tisíců. Výsledná teplota je závislá na výkonu laseru a ploše paprsku dopadající na vzorek.

Je zde zaneseno mnoho relativních chyb v průběhu výpočtu (měření). Chyba při každém měření v daném bodě vyplývající ze standardní odchylky, která se pohybovala okolo 10% – 15%. Dále také chyby Einsteinových koeficientův rozsahu až 30%. Tyto chyby se ve výsledku projeví v podobě desítek procent [8].

1.4 Rozšíření spektrálních čar

Fotony daných frekvencí jsou emitovány z důvodu rozdílnosti dvou specifických energetických hladin. Co tedy očekáváme je skokový nárůst spektrální čáry emitující na konkrétní frekvenci, totožná s Diracovou delta funkcí obrázek 7. Tato myšlenka vychází i z Bohrova modelu atomu. V praxi ale pozorujeme jev zvaný rozšíření spektrální čáry.

Získáme rozsah frekvencí. Daná hodnota intenzity nám určuje pravděpodobnost, na jaké frekvenci daný foton vyzařuje energii. Vhodné mechanismy pro rozšíření spektrální čáry pro emisi plazmatu jsou Dopplerovo a Starkovo rozšíření. Teorie použitá v této kapitole byla čerpána z [11], [13] a [12].



Obrázek 7: Diracova δ funkce, [18]

1.4.1 Dopplerovské rozšíření

Dopplerovské rozšíření je aplikováno v důsledku chaotického pohybu emitujících atomů. Jde o závislost na teplotě T a frekvenci f .

Jedná se o aplikaci Dopplerova jevu, který popisuje změnu frekvence a vlnové délky, přijímaného oproti vysílanému signálu. Tento jev nastává díky nenulovému pohybu vysílače a přijímače.

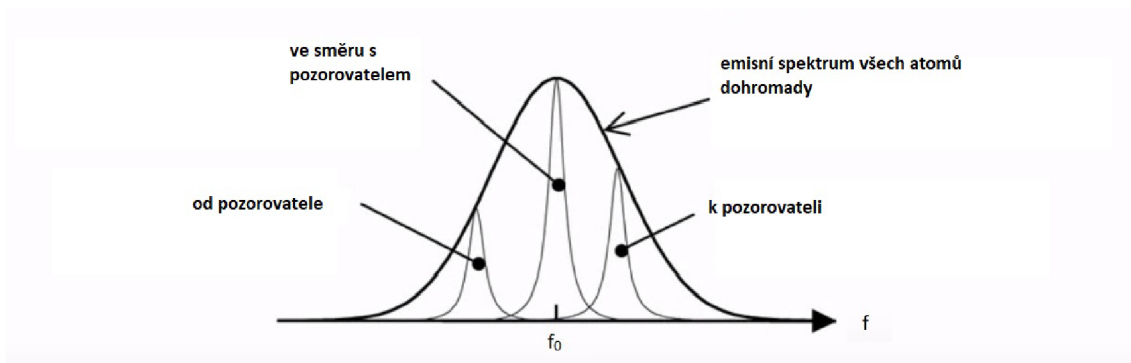
$$f = f_0 \left(1 \pm \frac{v}{c}\right), \quad (1.6)$$

kde zdroj vysílaného signálu je f_0 a pozorovatel jej přijímá jako f , v je rychlost pohybu vysílače či přijímače a c je rychlost vln v dané látce. Znaménko plus je použito při pohybu těles k sobě.

Mnoho atomů se pohybuje různým směrem a různou rychlostí. Díky tomu se může vyskytovat daná spektrální čára na určitém intervalu frekvencí. Dopplerovo rozšíření získáme, pomocí sestavení obálky těchto frekvencí s danou intenzitou. Tato obálka je charakterizována Gaussovou funkcí. Daný jev tedy můžeme aproximovat Gaussovým rozdělením:

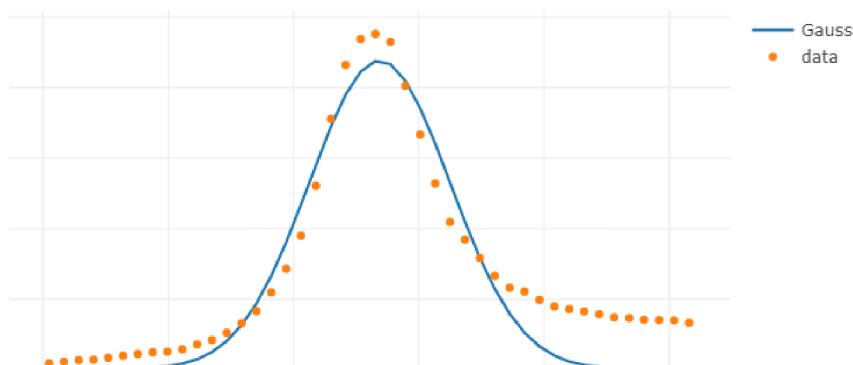
$$f(f) = A \exp\left(\frac{-(f - f_0)^2}{2\sigma^2}\right), \quad (1.7)$$

kde $\sigma = \sqrt{\frac{kT}{mc^2}} f_0$, k je Boltzmannova konstanta, m je hmotnost emitované částice, T je teplota a f_0 je frekvence na vrcholu funkce. Koeficient A je intenzitní faktor, amplituda intenzity. Dá se také brát jako normalizační podmínka pro kterou platí $A = \frac{1}{\sigma\sqrt{2\pi}}$ pro odpovídající Normální rozdělení $N(f_0, \sigma^2)$. Pro funkci platí, že integrál dané funkce $f(f)$



Obrázek 8: Na obrázku vidíme atomy pohybující se směrem od pozorovatele (nalevo od f_0) a atomy pohybující se ve správném směru k pozorovateli f_0 , ale také atomy mířící k pozorovateli (napravo od f_0) [37].

přes celý definiční obor je roven jedné. Jde tedy o vyjádření hustoty pravděpodobnosti. V této části vycházíme z [14].



Obrázek 9: Data jsou ze souboru hliníkového standardu (viz příloha H) Zdrojová data), ze kterých byla vytvořena charakteristická čára hliníku a vybrána jedna spektrální čára $\lambda = 394.459nm$ na níž byla aplikována Gaussova aproximace.

1.4.2 Starkovo rozšíření (tlakové rozšíření)

Vzniká díky vzájemné interakci nabitých částic a je úměrné počtu srážek za jednotku času. Elektrické pole generované elektrony a ionty v plazmatu narušuje energetické hladiny jednotlivých iontů, čímž rozšiřuje emisní linky [14]. Na dobře izolovanou čáru můžeme použít toto rozšíření. Tento děj se v praxi aproximuje Lorenzovou funkcí známou také jako Cauchy-Lorenzovo rozdělení. Odhad hustoty pravděpodobnosti emise na dané frekvenci

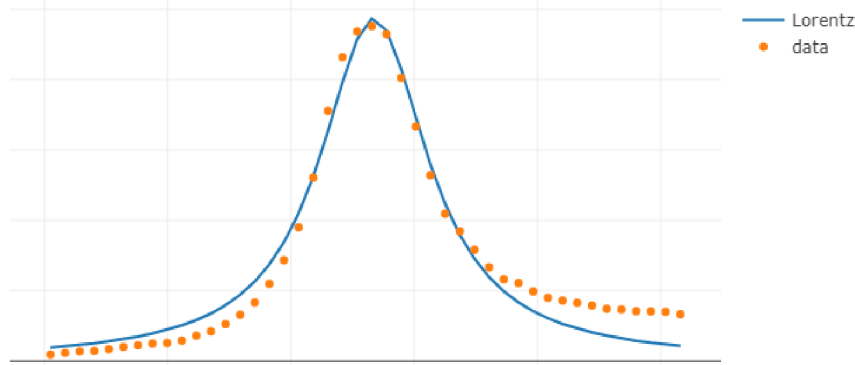
je možný, známe-li Starkův koeficient rozšíření λ .

$$g(f) = Af^2 \frac{\lambda}{(f - f_0) + \lambda}. \quad (1.8)$$

Parametr A je intenzitní faktor. Starkův koeficient rozšíření λ vyjádřen jako FWHM (full width at half maximum) v nanometrech, je dán rovnicí [20]

$$\lambda = 2w\left(\frac{n_e}{10^{16}}\right) + 3.5C\left(\frac{n_e}{10^{16}}\right)^{\frac{1}{4}}[1 - BN_D^{-\frac{1}{3}}]w\left(\frac{n_e}{10^{16}}\right) \quad (1.9)$$

V rovnici výše uvedené B je koeficient rovný 1.2 nebo 0.75 u iontových nebo neutrálních linek, w je parametr dopadu elektronu a A je iontové rozšíření. První člen pravé strany rovnice pochází z interakce elektronů. Druhý člen je generován interakcí iontů. $n_e(\text{cm}^{-3})$ je hustota elektronů. N_D je počet částic určité oblasti. Více viz [20]. Dané parametry potřebné k výpočtu jsou dohledatelné v literatuře od Griema [15].



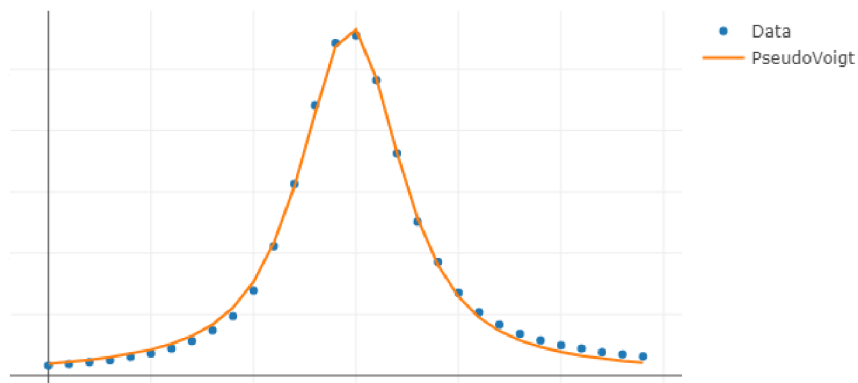
Obrázek 10: Data jsou ze souboru hliníkového standardu (viz příloha H) Zdrojová data), ze kterých byla vytvořena charakteristická čára hliníku a vybrána jedna spektrální čára $\lambda = 394.459\text{nm}$ na níž byla aplikována Lorentzova aproximace.

1.4.3 Voigtův profil

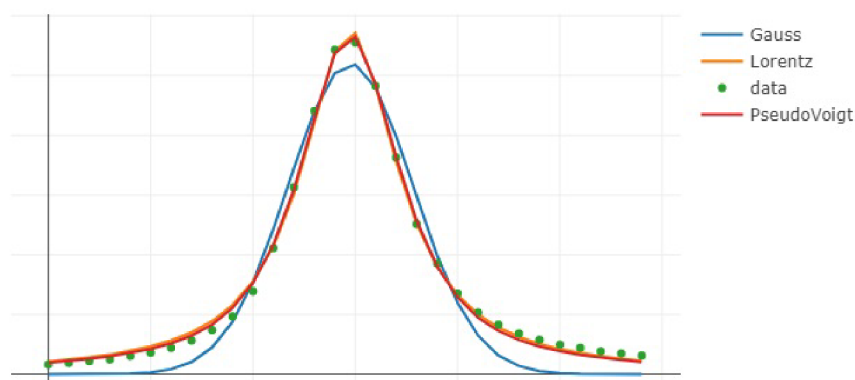
Diskrétní aproximaci Voigtova profilu nazýváme Pseudo Voigtův profil. Výsledný tvar spektrální čáry je nejlépe popsán Voightovým profilem, ve kterém je obsaženo jak Dopplerovo tak i Starkovo rozšíření. Jde o konvoluci těchto dvou aproximací [16]. Konvoluce vyjadřuje množství překrytí jedné funkce přes druhou.

$$(f * g)(f) = \int_{-\infty}^{\infty} f(\tau)g(f - \tau)d\tau, \quad (1.10)$$

kde g je Lorenzova funkce uvažována jako konvoluční jádro. U konvoluce jde o integraci součinu funkce f (Gaussovi) s jádrem g , parametr τ je integrační proměnná.



Obrázek 11: Data jsou ze souboru hliníkového standardu (viz příloha H) Zdrojová data), ze kterých byla vytvořena charakteristická čára hliníku a vybrána jedna spektrální čára $\lambda = 394.459nm$ na níž byla aplikována konvoluce Gaussovi a Lorenzovi funkce.



Obrázek 12: Data jsou ze souboru hliníkového standardu (viz příloha H) Zdrojová data), ze kterých byla vytvořena charakteristická čára hliníku a vybrána jedna spektrální čára $\lambda = 394.459nm$ na níž byly aplikovány různé aproximace.

2 Předzpracování dat

Data z reálného života získané měřeními fyzikálních veličin jsou citlivá na šum. Nekonzistentní data jsou typicky způsobena velkou velikostí dat. Nízká kvalita dat vede k nízké kvalitě konečných výsledků po zpracování. Otázkou je, jak data předzpracovat, aby se zlepšila jejich kvalita? Jednotlivé metody předzpracování dat se dají kombinovat. Některé postupy si ukážeme v této kapitole. Tyto postupy se dají použít pro sloupcové i řádkové úpravy. V kapitole předzpracování dat čerpáme primárně z poznatků z literatury [22], [31] a [32].

2.1 Detekce outlierů

Jako výstup z LIBS dostáváme matici statistických dat. Z každého měření získáme vlnové délky a jejich příslušné intenzity. Při zpracování těchto informací může nastat určitá anomálie měření, která výrazně vybočuje od ostatních. Toto měření je vhodné vyřadit z testovací množiny ještě před matematickým zpracováním dat. Jedná se o chybu, která by nám zkreslovala výsledek. Rozhodnutí zda se jedná o outlier či ne je subjektivní a musíme jej zhodnotit pomocí znalosti k dané problematice. Máme určité metody detekce, které nám toto rozhodování usnadní.

2.1.1 Detekce založená na blízkosti

Jde o metodu, kde si najdeme průměrný “nejtypičtější” bod z množiny dat. Pro vícerozměrná data, jejichž analýzou se zabýváme, budeme brát “nejtypičtější” bod jako vektor těžiště. Rovnice těžiště $X_T = (X_{T1}, X_{T2}, \dots, X_{Tm})$, kde rozměr matice je $X(n \times m)$

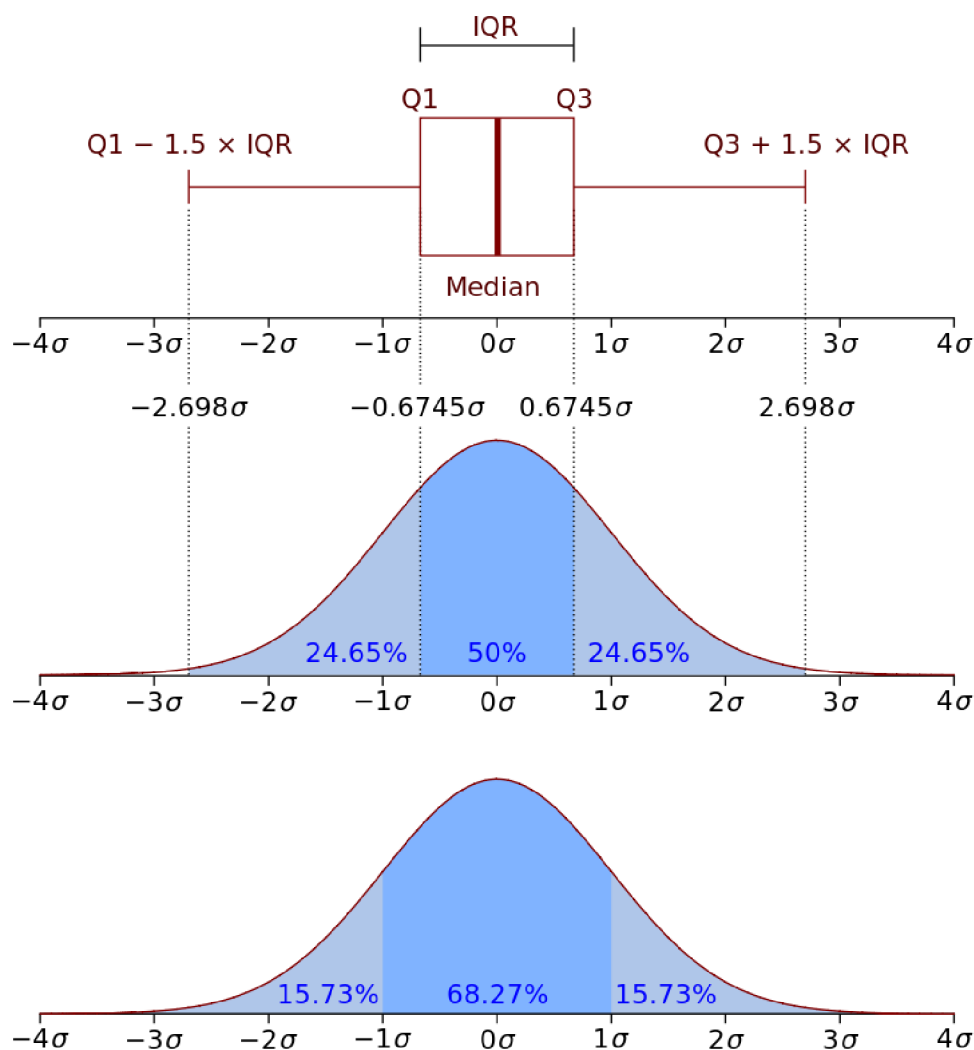
$$X_{Tj} = \frac{1}{n} \sum_{i=1}^n X_{ij}. \quad (2.1)$$

Pokud se dané měření výrazně odlišuje od tohoto těžiště, budeme dané měření brát jako outlier.

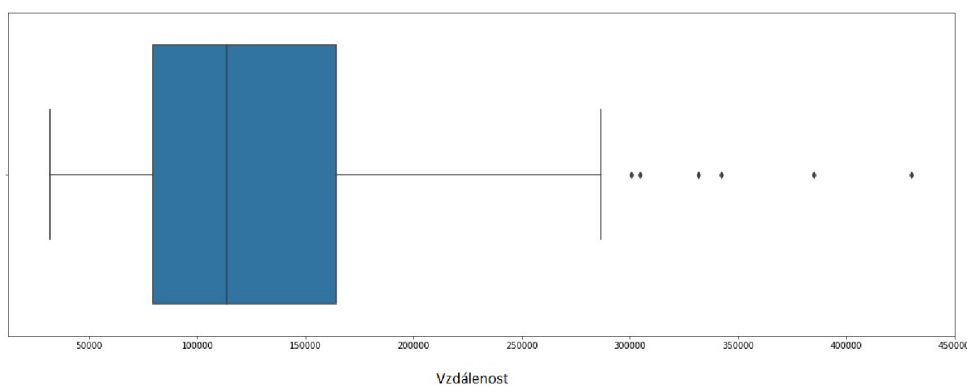
K detekci, jak moc se dané měření odlišuje od těžiště, můžeme využít metodu nejbližších sousedů. V Softwaru Python je tato metoda známá jako NearestNeighbors [33]. Výstupem z dané metody získáme pořadí vzorků a jejich vzdálenosti od těžiště. Vzdálenost může obecně představovat jakoukoliv metriku. My jsme použili Euklidovskou metriku, která se vypočítá v každé dimenzi a výsledná vzdálenost je suma přes všechny dimenze. Z metody se bohužel nedozvíme, jaká vzdálenost už je velká a měla by představovat hranici pro detekci outlierů.

Chceme-li získat rovnou výrok, zda jde u nějakého měření o outlier, můžeme si pomoci statistikou a kvantily. Budeme předpokládat Normální rozdělení. Zvolíme si rozpětí intervalu podle procent dat, které chceme zachovat. Například interkvartilové rozpětí, které udává interval mezi 25% percentilem a 75% percentilem statistického souboru. Nebo také pomocí směrodatných odchylek. Vybereme-li například $\pm 3\sigma$ (směrodatná odchylka), tak tento údaj pokrývá 99,73% statistického souboru. Tato klasifikace pomocí směrodatných odchylek je známá také jako z-score. Toto rozložení má vždy průměr nulu, směrodatnou odchylku jedna a je interpretováno jako počet směrodatných odchylek nad nebo pod průměrem.

Měření, která neleží v tomto intervalu jsou hodnocena jakou outliersy.



Obrázek 13: Na obrázku vidíme interkvartilové rozpětí IQR, které popisuje 50% statistického souboru. Dále $1,5 \times IQR$, které popisuje asi 99,3% statistického souboru, a nakonec výběr podle směrodatné odchylky $\pm 3\sigma$, které popisuje 99,73% statistického souboru, zdroj: [34].



Obrázek 14: Diagram znázorňující 6 outlierů a interkvartilové rozpětí, data jsou ze souboru hliníkového standardu (viz příloha H) Zdrojová data): 100 měření spektrálních čar hliníkového vzorku.

Na obrázku 14 je znázorněn box plot. Modrá oblast je IQR oblast, v níž je černým páskem znázorněna střední hodnota. Dále mezi zbývajícími dvěma svislými čarami je oblast $1,5 \times \text{IQR}$ a dále mimo interval jsou zobrazeny body, které bychom pomocí této analýzy nazvali za outliery a odstranili ze statistického souboru.

2.2 Redukce dimenze

Dimenzi můžeme redukovat výběrem konkrétních významných čar. Postupujeme následovně: ke každé čáře najdeme její okolí, na kterém provedeme Voigtovu aproximaci. Nahradíme okolí čáry pouze jejím vrcholem a hodnotu intenzity spočítáme jako plochu pod Voigtovou funkcí. Tímto jsme schopni celé široké spektrum nahradit pouze pár hodnotami. Teoreticky je možné vybrat pouze jednu takovou čáru pro každý prvek vyskytující se ve vzorku. Jde o subjektivní výběr čar pracovníka. Dále můžeme dimenzi redukovat pomocí metody hlavních komponent (PCA), která je detailněji rozebrána v pozdější kapitole.

2.3 Scaling

Scaling neboli standardizace odstraňuje zkreslení zapříčiněné nerovnoměrným rozdělením daných proměnných. Při standardizaci dat převádíme data z různých zdrojů či různých formátů do jednoho standardního formátu.

2.3.1 Centrování zdrojové matice X

Jde o odstranění závislosti na jednotkách a na parametru polohy, tedy aby průměr každé dimenze byl 0. Toho dosáhneme odečtením průměru každé dimenze od každého prvku této dimenze. Pokud máme tedy zdrojovou matici o m dimenzích, musíme od každého prvku jednotlivé dimenze X_{ij} odečíst průměr dané dimenze $\bar{X}_j = \sum_{i=1}^n \frac{X_{ij}}{n}$. Tedy centrovaná data dané dimenze jsou

$$Xc_{ij} = X_{ij} - \bar{X}_j, \quad (2.2)$$

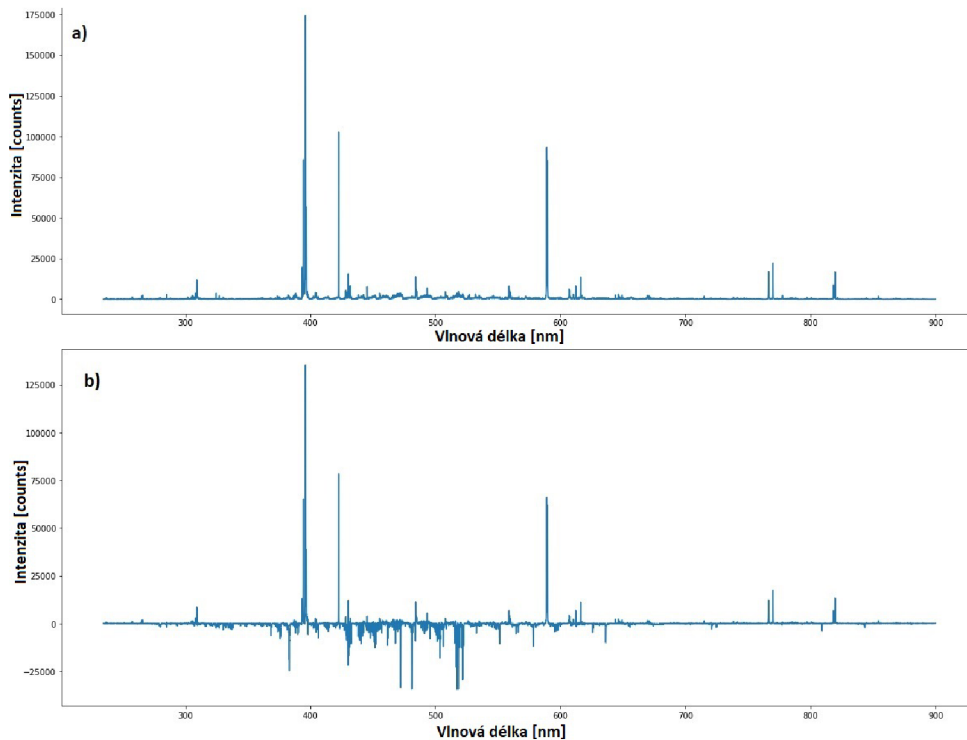
kde Xc můžeme nazvat sloupcovým vektorem a také interpretovat jako $Xc = (X_1 - \mu_1, X_2 - \mu_2, \dots, x_m - \mu_m)^T$, μ je střední hodnota dané dimenze.

2.3.2 Standardizace směrodatnou odchylkou

Hodnotu po standardizaci získáme odečtením aritmetického průměru dané proměnné (sloupce) od původní hodnoty a vydělené směrodatnou odchylkou dané proměnné (sloupce)

$$X_s = \frac{X - \bar{X}}{\sigma}. \quad (2.3)$$

Nová hodnota X_s má nulový průměr a rozptyl roven jedné. Tento typ zpracování dat není vhodný na data, jenž jsou zatížena šumem. Šum se umocní a znehodnotí nám zdrojová data. Také není vhodné metodu používat v případech kdy dané rozdělení není normální a také při výskytu odlehlých bodů v datech. Metodu je tedy vhodné použít jako druhý krok po detekci a odfiltrování odlehlých bodů (outlierů).



Obrázek 15: Na obrázku je vybráno pouze jedno měření, na kterém jsou: a) surová naměřená data, b) data po vycentrování přes sloupce. Jde o zdrojovou matici dat (viz příloha H) Zdrojová data), z níž jsou na ukázkou vybrány hodnoty druhého měření.

2.3.3 MinMax standardizace

Tento typ standardizace je vhodný použít u proměnných, které mají velké rozdíly mezi hodnotami, nemají normální rozdělení nebo obsahují odlehlé body (outliery). Pevný rozsah hodnot metody od 0 do 1. Metoda má menší rozsah, znamenající menší standardizované odchylky. Tento jev se pozitivně projeví k potlačení vlivu outlierů.

$$X_{mm} = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (2.4)$$

2.4 Odečtení pozadí

Spektra získána pomocí laserem buzeného plazmatu obsahují cennou část signálu, ale také nežádoucí signál pozadí, který se snažíme zredukovat. Numerická oprava signálu může vést ke ztrátě informace.

Toto pozadí vzniká při měření díky například: “dark current” u detektoru, což je elektrický proud objevující se na CCD čipu vznikající jinak než dopadem fotonů a šumu z laserového pulzu z kontinuálního záření volných elektronů. Tento jev je závislý na nastavení doby snímání tzv. gate delay.

Jedním ze způsobů jak vyhladit a zbavit šumu data je vlnková transformace. Tímto postupem se v diplomové práci zabývat nebudeme, pro detailnější vhléd do této problematiky viz [35].

3 PCA

V oblasti zabývající se LIBS se počet vědeckých článků na téma PCA stále zvyšuje. Z těchto článků plyne PCA jako jedna z primárních metod vhodná jak pro vizualizaci dat, tak jejich následnou analýzu. Jedním z takových článků je článek pana doktora Pořízky a spol. [6].

Metoda analýzy hlavních komponent (principal component analysis-PCA) patří mezi nejstarší a nejpoužívanější metody pro zpracování vícerozměrných dat. Byla odvozena již v roce 1901 Pearsonem.

Tuto metodu používáme, potřebujeme-li nějak vizualizovat data a získat o nich lepší představu. Cílem je získat vzájemnou korelaci původních dat, tedy zjistit důležité informace z původní zdrojové matice a následně je převést do nových ortogonálních proměnných. Při vizualizaci dat narážíme na problém znázornění dat. Chceme-li znázornit p dimenzionální prostor pomocí 2D projekce, bude zapotřebí znázornit $\frac{1}{2}p(p-1)$ dílčích 2D grafů. V našem případě chceme znázornit objekty v závislosti na jejich znacích. V tom nám pomůže PCA, která se používá k redukcí vícerozměrných dat a také k potlačení šumu ve zdrojové matici. Metoda analýzy hlavních komponent nám pomůže roztřídit v prostoru objekty, které mají nějakou společnou charakteristickou vlastnost (znak), či kombinaci vlastností (znaků). Vycházíme primárně z [3] a [22].

3.1 Obecně

Metoda je postavena na zjednodušení popisu skupiny znaků, které jsou vzájemně lineárně závislé (korelované). Jde o rozklad *zdrojové matice* dat do *matice strukturní* a *matice šumové*. Jedná se v podstatě o lineární transformaci původních znaků na nové nezávislé (nekorelované) proměnné, které budeme nazývat *hlavní komponenty*, nebo také latentní proměnné.

Nově získané proměnné obsahují lineární kombinace původních znaků. Primární charakteristikou hlavních komponent je jejich rozptyl (variabilita). Ta nám udává jaké množství informace nese daná komponenta o celé skupině zdrojových dat. Hlavní komponenty dále řadíme dle klesajícího rozptylu. První hlavní komponenta je charakterizována největším rozptylem a tedy obsahuje nejvíce informace a poslední hlavní komponenta obsahuje naopak informace nejméně.

Máme-li znak s malým rozptylem, tak nám nepomůže v rozlišování mezi objekty. Analýza hlavních komponent se používá ke snížení dimenze dat, tedy ke snížení počtu znaků s malým rozptylem, které nám výrazně nepřispívají k rozlišení objektů. Vypuštěním znaků s malým rozptylem znamená snížení dimenze úlohy bez velké ztráty informace. Použijeme-li prvních několik hlavních komponent, pak toto snížení dimenze úlohy neznámá snížení počtu původních znaků, jelikož každá hlavní komponenta je lineární kombinací předešlých charakteristických znaků.

Hlavní výhodou metody je, že vyšetřujeme pouze "malý" počet nekorelovaných hlavních komponent namísto vyšetřování velkého počtu původních znaků s vazbami mezi sebou. Kapitola vychází primárně z [5], [7] a [22].

3.2 Matematický popis

Mějme zdrojovou matici dat $X(n \times m)$ obsahující n objektů a m znaků. Objekty n jsou pozorování a znaky m jsou proměnné, tedy nějaké vlastnosti daných objektů. Každý znak je znám pro všechna pozorování. Výhodou PCA je to, že lze použít libovolný počet proměnných ve zdrojové matici X k vícerozměrné charakterizaci. To můžeme provést z důvodu, že každá hlavní komponenta je lineární kombinací původních znaků. Podstata hlavní myšlenky je ve sledování rozptylu.

Provedeme transformaci původních znaků x_j , $j = 1, \dots, m$ do menšího počtu latentních proměnných y_j . Těchto nových latentních proměnných (hlavních komponent) je méně. Vystihují téměř celou proměnlivost znaků a jsou vzájemně nekorelované.

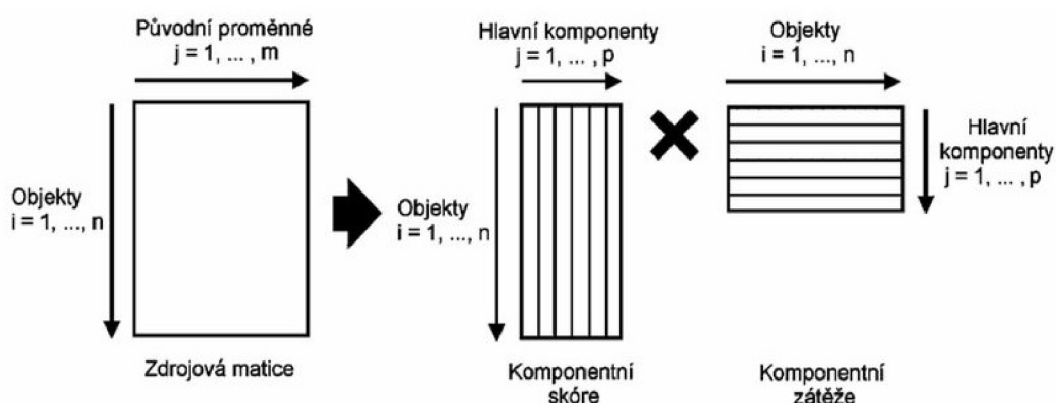


Schéma maticového výpočtu v PCA, zdroj [22] str.70.

Počet užitých hlavních komponent budeme značit k . Na levé straně maticového schématu můžeme vidět zdrojovou centrovanou matici X_c . Na pravé straně můžeme vidět rozklad matice X_c na matici komponentních skóre $T(n \times k)$ a matici komponentní zátěže (komponentních vah) $V^T(k \times m)$. Kapitola vychází primárně z [3] a [4].

3.3 Získání hlavních komponent

První hlavní komponenta y_1 je lineární kombinací vstupních znaků, která popisuje největší rozptyl mezi všemi ostatními lineárními kombinacemi

$$y_1 = \sum_{j=1}^m V_{1j} x_{cj}. \quad (3.1)$$

Pro získání hlavních komponent je potřeba znát kovarianční matici. Kovarianční matici můžeme popsat jako míru vzájemné vazby mezi dvěma dimenzemi. Obecný vzorec pro kovarianci

$$Cov(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n - 1)}, \quad (3.2)$$

X a Y jsou jednotlivé dimenze a n je počet jejich prvků. Výsledek kovariance musí ležet v intervalu od -1 do 1 . Kladný výsledek značí kladnou vazbu mezi prvky jednotlivých dimenzí, tedy rostou-li hodnoty v jedné dimenzi, budou růst i pro druhou. Analogicky pro záporný výsledek. Výsledek kolem 0 mezi daty je buď velmi malá nebo žádná závislost. Kovariance jen jedné dimenze $X = Y$ je rozptyl. Kovarianční matici, kterou budeme dále potřebovat, je nutné vypočítat pro naše m rozměrná data. Musíme tedy spočítat kovarianci mezi všemi dimenzemi. Získáme matici o rozměrech $(m \times m)$. Například matice pro 3 dimenze by vypadala následovně

$$C = \begin{pmatrix} Cov(m_1, m_1) & Cov(m_1, m_2) & Cov(m_1, m_3) \\ Cov(m_2, m_1) & Cov(m_2, m_2) & Cov(m_2, m_3) \\ Cov(m_3, m_1) & Cov(m_3, m_2) & Cov(m_3, m_3) \end{pmatrix}.$$

Výsledná matice je symetrická kolem hlavní diagonály a na hlavní diagonále obsahuje rozptyly jednotlivých dimenzí.

Jak bylo výše uvedeno chceme maximalizovat rozptyl

$$D(y_1) = V_1^T C V_1, \quad (3.3)$$

kteřý je závislý na velikosti vektoru V_1 o rozměru $(m \times 1)$. C charakterizuje kovarianční matici. Dále je třeba zavést omezení pro vektor V_1 . Omezení které zavedeme se nazývá normalizace $V_1^T V_1 = 1$. Pro komponenty vyšších řádů přibude ještě další omezení, a to $V_j^T V_j = 1 \wedge V_i^T V_j = 0$, pro všechna $i \neq j$, tato druhá podmínka zajišťuje kolmost hlavních komponent.

Pro výpočet vhodných vlastních vektorů V_1, V_2, \dots, V_m budeme muset řešit sérii maximalizačních úloh s omezením. Jedná se o statistickou optimalizaci funkcí více proměnných a problém vede na řešení pomocí *Lagrangeových multiplikátorů*. Ukážeme si řešení pro první hlavní komponentu. Toto odvození je primárně přejato z [4].

Mějme tedy funkci $D(y_1) = V_1^T C V_1$ a omezení $V_1^T V_1 = 1$.

Vytvoříme Lagrangian

$$L_1 = V_1^T C V_1 - l_1(V_1^T V_1 - 1), \quad (3.4)$$

l_1 je Lagrangeův multiplikátor a také polynom m -tého stupně s kořeny $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$. Rovnici (3.4) můžeme přepsat pro lepší orientaci do tvaru

$$L_1 = \sum_{i=1}^m \sum_{j=1}^m V_{1i} C_{ij} V_{1j} - l_1 \left(\sum_{i=1}^m V_{1i}^2 - 1 \right). \quad (3.5)$$

Podmínka pro maximalizaci V_1 je derivace Lagrangianu podle V_1

$$\frac{\partial L_1}{\partial V_1} = 2CV_1 - 2l_1V_1 = (C - l_1E) V_1 = 0. \quad (3.6)$$

Víme, že $V_1 \neq 0$ a tedy $(C - l_1E)$ musí být singulární

$$\det(C - l_1E) = 0. \quad (3.7)$$

Po úpravě poslední rovnosti (3.6) a vynásobením V_1^T získáme

$$V_1^T C V_1 = l_1 V_1^T V_1 = l_1. \quad (3.8)$$

Vypočteme l_1 vektor a rozptyl bude maximální pro vlastní číslo $l_{1\lambda} = \max(\lambda_1, \lambda_2, \dots, \lambda_m)$.
dále dosadíme do rovnice (3.8) $l_{1\lambda}$ za l_1

$$V_1^T C V_1 = l_{1\lambda} \quad (3.9)$$

a dopočítáme vlastní vektor V_1 pro první hlavní komponentu. Pro následující komponenty řešíme maximalizační úlohu pro Lagrangián

$$L_j = V_j^T C V_j - l_j(V_j^T V_j - 1) - 2 \sum_{i=1}^{j-1} l_{ci}, \quad (3.10)$$

l_j a l_{ci} jsou Lagrangeovy koeficienty, položením rovnice rovno nule a postupným vynásobením členy V_i , $i = 1, \dots, j-1$ a po úpravě nám výjde $l_{ci} = 0$ pro $i = 1, \dots, j-1$, že $l_{j\lambda} = \max(\lambda_1, \lambda_2, \lambda_{m-j-1})$, tedy rovno nejvyššímu zbylému kořenu rovnice

$$\det(C - l_j E) = 0. \quad (3.11)$$

Podmínka maxima se nám zjednodušila na tuto rovnici. V_j značí j -tý vlastní vektor odpovídající kořenu $l_{j\lambda}$. Došli jsme tedy k zjištění, že vektor V_j je vlastní vektor kovarianční matice C , kterému odpovídá j -té největší vlastní číslo.

Můžeme tedy kovarianční matici rozepsat ve tvaru

$$C = V D V^T. \quad (3.12)$$

Matice V ($m \times m$) obsahuje ve sloupcích vektory V_j a je také ortogonální $V V^T = E$, E jednotková matice. D je diagonální matice ($m \times m$), která obsahuje na diagonále vlastní čísla $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ kovarianční matice. Z rovnice (3.12) vidíme, že rozptyl $D(y_j) = \lambda_j$.

Pro analýzu dat budeme také potřebovat *komponentní skóre objektů*. Popisuje, jak moc jsou dané původní znaky obsaženy v příslušné hlavní komponentě.

Máme-li k dispozici hodnoty m znaků pro libovolný i -tý objekt $x_i = (x_{i1}, \dots, x_{im})^T$, známe také vektor středních hodnot $\mu = (\mu_1, \mu_2, \dots, \mu_m)^T$ a vlastní vektory V_j , $j = 1, \dots, m$, jsme schopni určit složku komponentního skóre objektu $t_i = (t_{1i}, \dots, t_{mi})^T$

$$t_{1i} = V_1^T (x_i - \mu), \quad t_{ji} = V_j^T (x_i - \mu), \quad j = 2, \dots, m. \quad (3.13)$$

Pro n objektů získáme *matici komponentního skóre*

$$T = \begin{pmatrix} t_1^T \\ \dots \\ t_n^T \end{pmatrix}.$$

3.3.1 Alternativní zavedení optimalizačního problému

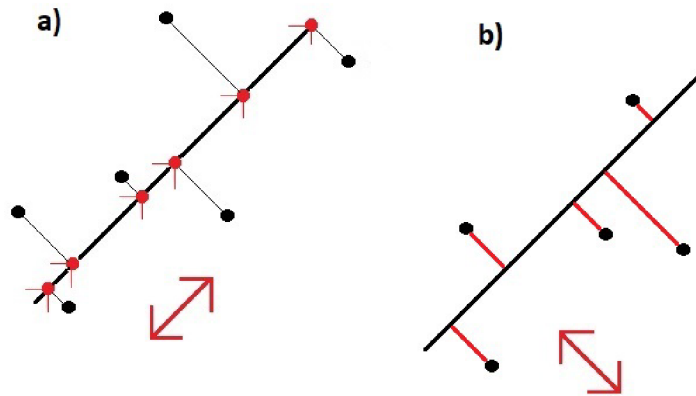
Optimalizační problém můžeme zavést alternativním způsobem. Výše uvedená metoda optimalizace je pomocí maximalizace rozptylu dat, jakmile jsou promítnuty do nového prostoru. Tato alternativa převádí problém na minimalizační úlohu. Minimalizuje vzdálenost projekce ve smyslu nejmenších reziduálních čtverců a je ekvivalentní k výše uvedenému

postupu[19].

Můžeme tedy psát jako minimalizační úlohu

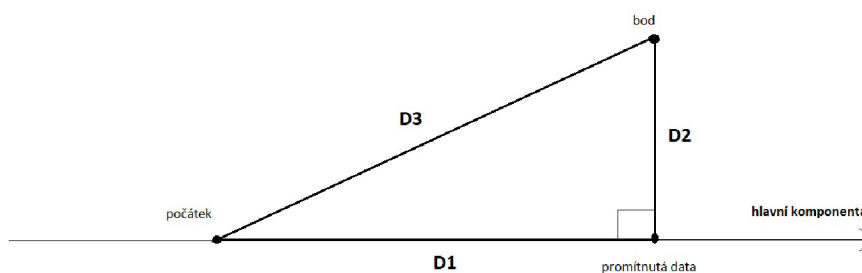
$$\|X_c - T V_k\|^2, \quad (3.14)$$

s omezením $V_k^T V_k = E$, E je jednotková matice. Bereme k z m hlavních komponent, kde $k < m$ a $\|\cdot\|^2$ je Euklidovská norma.



Obrázek 16: Vizualizace alternativního zavedení a ukázka jaké vzdálenosti v daných případech: a) maximalizujeme rozptyl, b) minimalizujeme rezidua.

Ukázka ekvivalence mezi minimalizací čtverců reziduí a maximalizací rozptylu. Vezmeme-li jeden vzorek a_i jako řádek X_c . Pak tento bod se projeví v rozptylu jako $Cov(a_i, a_i)$ což je vlastně Euklidovská vzdálenost $\|a_i\|^2$. Pomocí Pythagorovi věty ukážeme, že tento celý rozptyl je roven součtu ztrát rozptylu (čtvercová rezidua) a zbývajících rozptylu. Na obrázku 17 tohle můžeme vidět ve dvourozměrném prostoru. D_3 je počáteční rozptyl, D_2



Obrázek 17: Vizualizace alternativního zavedení. D_3 je počáteční rozptyl, D_2 je zbývajících rozptyl a D_1 je ztráta rozptylu.

je zbývajících rozptyl a D_1 je ztráta rozptylu.

$$D_3^2 = D_1^2 + D_2^2 \quad (3.15)$$

$$\|a_i\|^2 = \|T_i V\|^2 + \|a_i - T_i V\|^2. \quad (3.16)$$

Pro zachování rovnosti z Pythagorovi věty $\|a_i\|^2$ je konstanta a máme na výběr buď maximalizovat člen $T_i V$ nebo minimalizovat člen $a_i - T_i V$.

Je tedy ekvivalentní maximalizovat maximalizovat rozptyl nebo minimalizovat ztrátu rozptylu (čtvercová rezidua) k nalezení hlavních komponent. V kapitole vycházíme z [4] a [19].

3.4 Predikce na základě modelu

Důležitou interpretací PCA je na jejím základě vytvořit model.

Rozdílem mezi souřadnicemi objektů v původních znacích a pomocí hlavních komponent vzniká ztráta informace díky projekci do menší prostorové dimenze. Tato ztráta se nazývá mírou těsnosti proložení modelu PCA [4].

Při rekonstrukci obecně používáme k z m hlavních komponent. Projeví se ztráta informace vznikem chybové matice O ($n \times m$). Pro vznik chybové matice musí platit $k < m$, platí tedy

$$X_c = T V_k + O = \hat{X}_c + O. \quad (3.17)$$

\hat{X}_c ($n \times k$) je predikce matice vstupních dat X_c ($n \times m$). Při vytvoření této predikce musíme počítat se vznikem chybové matice způsobené snížením dimenze úlohy. O je chybová matice, kterou můžeme vyčíslit jako

$$O = X_c - \hat{X}_c. \quad (3.18)$$

Nabízí se otázka, jak dobře daný model vystihuje skutečnost? K hodnocení kvality modelu se snažíme vyjádřit určitou podobnost mezi maticemi X_c a \hat{X}_c . Hledáme tedy společné struktury mezi dvěma datovými soubory popisující stejné objekty. Jedná se o bilineární regresní model. Je tedy vhodné zmínit dva přístupy, které nám otázku pomůžou zodpovědět.

Reziduální součet čtverců (RESS)

$$RESS = \left\| X_c - \hat{X}_c \right\|^2 = \|O\|^2 = \text{Tr}(O^T O), \quad (3.19)$$

kde $\|A\| = \sqrt{\sum_{i=1}^n \sum_{j=1}^m a_{i,j}^2}$ je Euklidovská norma (tedy odmocnina součtů všech čtverců elementů matice A), $\text{Tr}(A) = \sum_{j=1}^m a_{j,j}$ je stopa matice (tedy součet všech diagonálních prvků matice A), kde matice A je vzata pouze na vysvětlenou pomocných definic.

Čím menší je hodnota tím lepší je PCA model [19].

RV koeficienty

Jedná se o zobecněný Pearsonův koeficient pro vícerozměrný problém. Měří blízkost dvou matic \hat{X}_c ($n \times k$) a X_c ($n \times m$), které jsou obdélníkové. Mají stejný počet objektů (řádků) a různý počet znaků (sloupců). Sloupce musí být centrované, jejich střední hodnota je rovna nule.

První spočítáme pomocné matice $G = X_c X_c^T$ a $H = \hat{X}_c \hat{X}_c^T$. Obě matice jsou pozitivně semidefinitní a mají stejnou dimenzi n .

$$RV = \frac{\text{Tr}(G^T H)}{\sqrt{\text{Tr}(G^T G) \text{Tr}(H^T H)}} \quad (3.20)$$

$$RV = \frac{\sum_{i=1} \sum_{j=1} g_{i,j} h_{i,j}}{\sqrt{\sum_{i=1} \sum_{j=1} g_{i,j}^2 \sum_{i=1} \sum_{j=1} h_{i,j}^2}}. \quad (3.21)$$

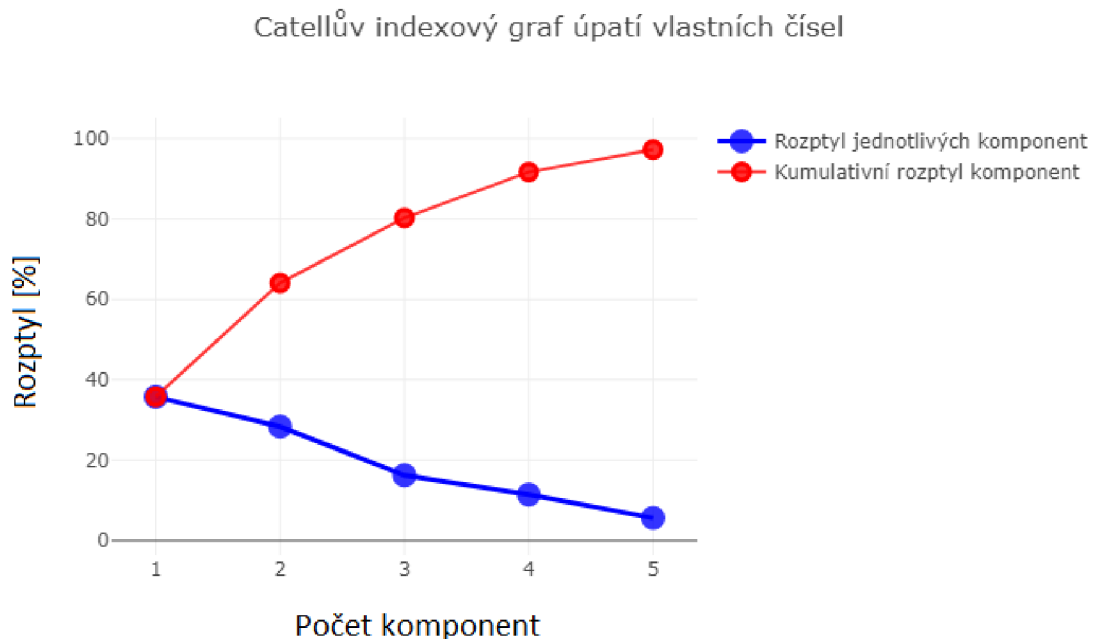
Tento koeficient nabývá hodnoty mezi 0 a 1. Čím blíž 1, tím jsou si matice podobnější.

3.5 Grafická reprezentace výstupu PCA

3.5.1 Získání počtu hlavních komponent

Jedno z možných kritérií k výběru počtu hlavních komponent je **Kaiserovo kritérium**, jedná se pouze o analytické porovnávací kritérium. To doporučuje dále nepokračovat v hledání dalších hlavních komponent s vlastním číslem menším než 1. Vyšší hlavní komponenty a jejich váhy mohou popisovat pouze šum vstupní matice, je tedy velmi důležité správně rozhodnout o počtu hlavních komponent. Dalším kritériem pro výběr vhodného počtu hlavních komponent je **Cattelův indexový graf úpatí vlastních čísel (Scree Plot)**. Jedná se o diagram vlastních čísel $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ zdrojové matice X v závislosti na indexu i vlastního čísla. Zobrazuje relativní velikost jednotlivých vlastních čísel. Lze podle něj určit počet k významných komponent. Vlastní čísla jsou seřazeny sestupně podle velikosti a hranice mezi významnými a nevýznamnými hlavními komponentami je znázorněna zřetelným zlomovým místem. Vybereme k -tici komponent, které objasňují zadané procento znaků. Typicky se volí hodnota $P = 70 - 90 \%$. Vybíráme k takové, že platí

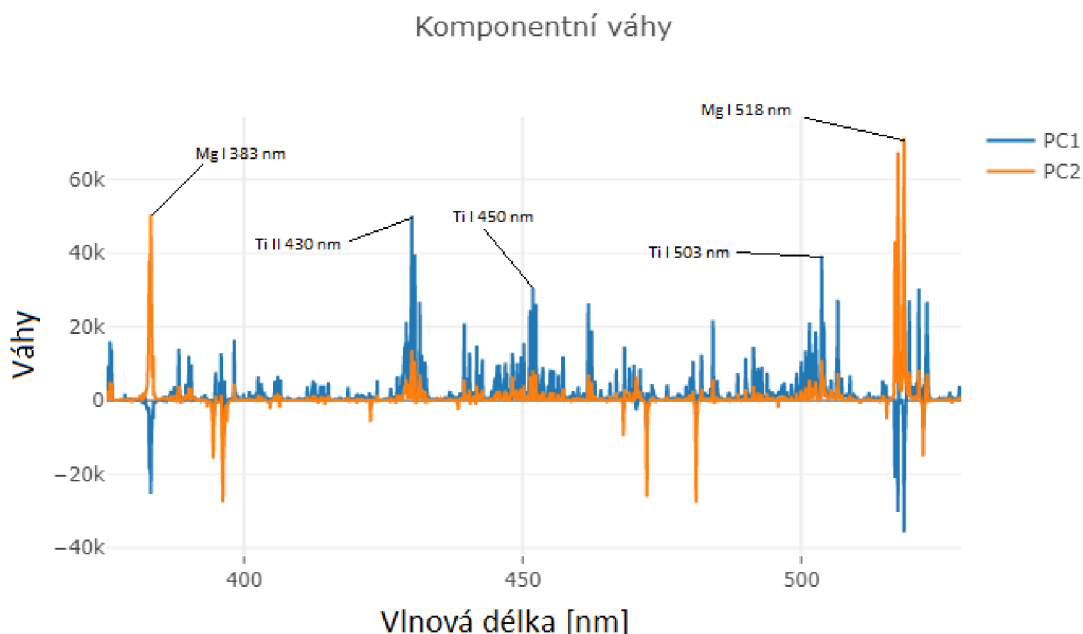
$$100 \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^m \lambda_i} \geq P. \quad (3.22)$$



Obrázek 18: Cattelův indexový graf úpatí vlastních čísel. Jde o zdrojovou matici dat (viz příloha H) Zdrojová data).

3.5.2 Graf komponentních vah (Plot Components Weights)

Zobrazuje komponentní váhy pro zvolené hlavní komponenty. Body v grafu znázorňují konkrétní znaky a v grafu nás zajímá vzdálenost jednotlivých bodů (znaků). Když je vzdálenost dvou bodů malá, jde o silnou korelaci mezi danými body. Můžeme nalézt i určité shluky daných znaků, které spolu vzájemně korelují. Graf nám ukazuje jakou měrou přispívají jednotlivé původní znaky do hlavních komponent. Znaky s vysokou proměnlivostí v objektech mají vysoké hodnoty komponentních vah, tyto znaky v diagramu leží daleko od počátku. Znaky ležící blízko počátku mají malou důležitost. Pro velké množství proměnných zobrazujeme pouze vztah mezi původními proměnnými a danou hlavní komponentou, z důvodu přehlednosti (obrázek 19).



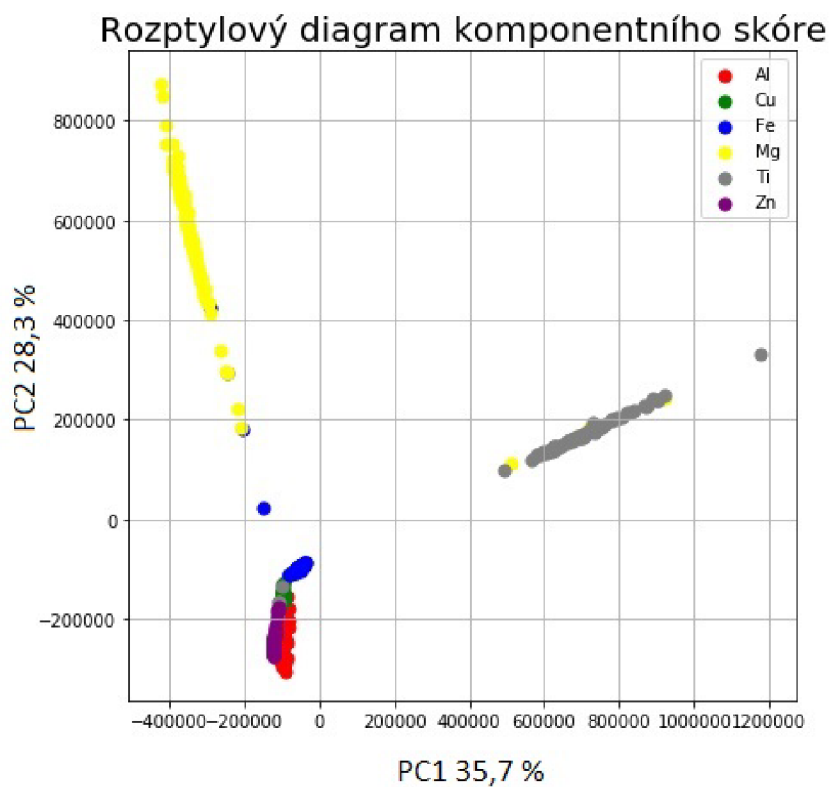
Obrázek 19: Graf komponentních vah pro první a druhou hlavní komponentu, zdrojová matice dat (viz příloha H) Zdrojová data).

3.5.3 Rozptylový diagram komponentního skóre (Scatterplot)

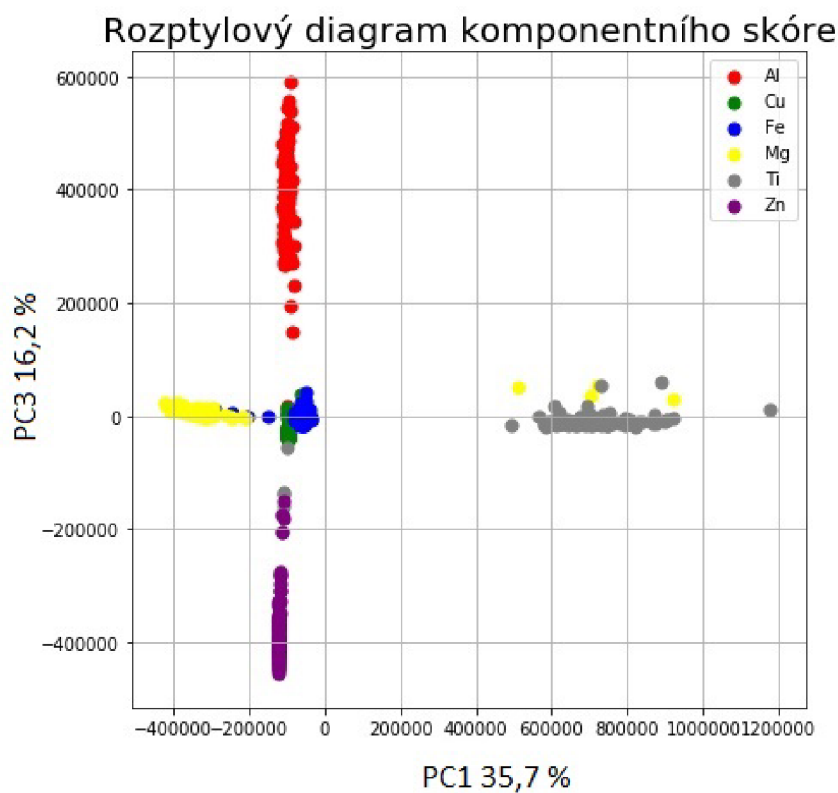
Zobrazuje komponentní skóre daných objektů pro zvolené hlavní komponenty, obvykle první dvě hlavní komponenty, tedy t_1 , t_2 . Body v tomto grafu jsou t_{1i}, t_{2i} , kde $i = 1, \dots, n$. Objekty dokonale rozmístěny v prostoru znamenají dokonalé rozlišení daných objektů. Můžeme také nalézt shluky objektů a také objekty výrazně se lišící od ostatních objektů. Objekty blízko sebe si jsou podobné.

Na obrázku 20 tedy můžeme vidět význam komponentních vah při vizualizaci komponentního skóre. Tyto váhy nám kladly význam na určité spektrální čáry. Konkrétně na čáry titanu v první hlavní komponentě a na čáry hořčíku ve druhé hlavní komponentě. Z grafické reprezentace PCA vidíme, jak se nám data rozdělila v novém prostoru a vytvořila shluky titanu a hořčíku, jak jsme již předpokládali z komponentních vah.

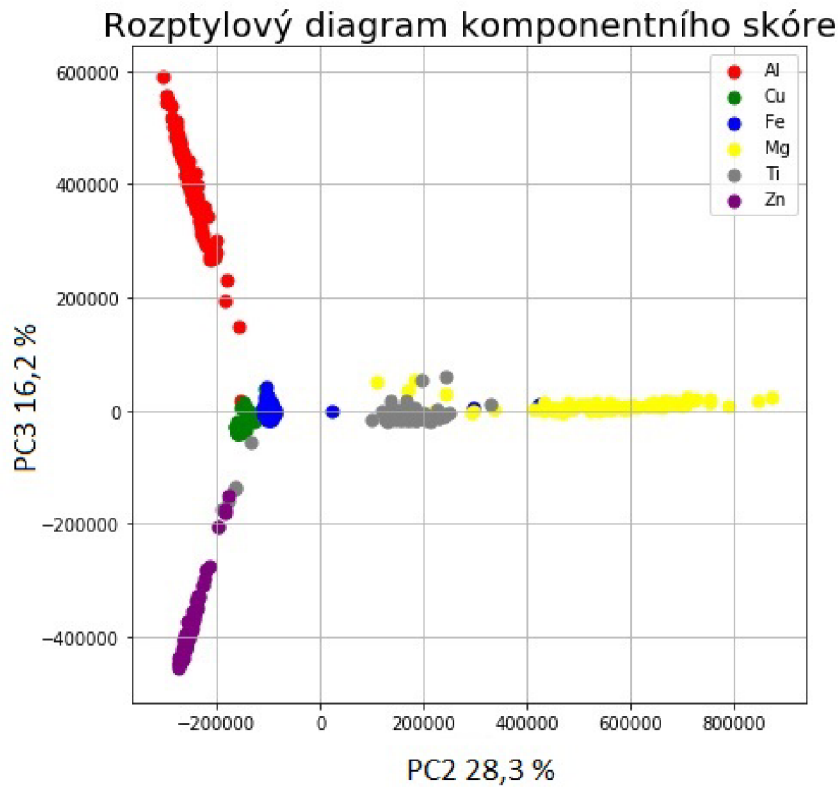
Jedním z možných dalších přístupů by bylo odstranění vzorků, nacházející se v oddělených



Obrázek 20: Diagram komponentního skóre pro první dvě hlavní komponenty PC1 x PC2.



Obrázek 21: Diagram komponentního skóre pro dvě hlavní komponenty PC1 x PC3.



Obrázek 22: Diagram komponentního skóre pro dvě hlavní komponenty PC2 x PC3.

oblastech, které jsme schopni klasifikovat (Ti, Mg, Al, Zn) a dále znova aplikovat celý postup PCA na zmenšenou matici neobsahující vzorky z těchto již námi známých oblastí.

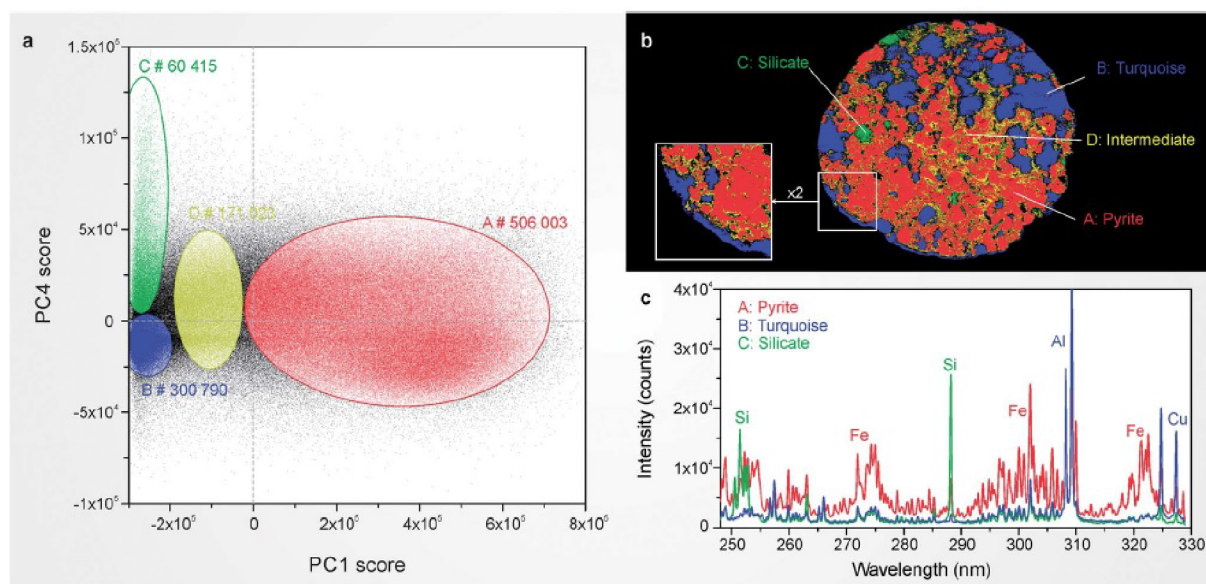
3.5.4 Dvojný graf(Biplot)

Jedná se o kombinaci předchozích dvou grafů. Body zde indikují objekty a vektory jsou projekce znaků. Úhel mezi průvodiči dvou znaků popisuje nepřímou úměrnost korelace daných znaků, čím menší úhel, tím větší korelace. Každý průvodič má své souřadnice na daných hlavních komponentách a projekce tohoto průvodiče na dané hlavní osy je úměrná velikosti příspěvku dané hlavní komponentě. Když se objekt nachází v blízkosti určitého znaku, znamená to, že tento objekt obsahuje velký podíl právě tohoto znaku. Všechny metody grafické reprezentace vychází primárně z [2], [3], [5] a [6].

4 Praktická část

V praktické části využijeme teoretických podkladů k analýze a interpretaci optických emisních spekter definovaných výše. Jedná se o proces laserem buzeného plazmatu.

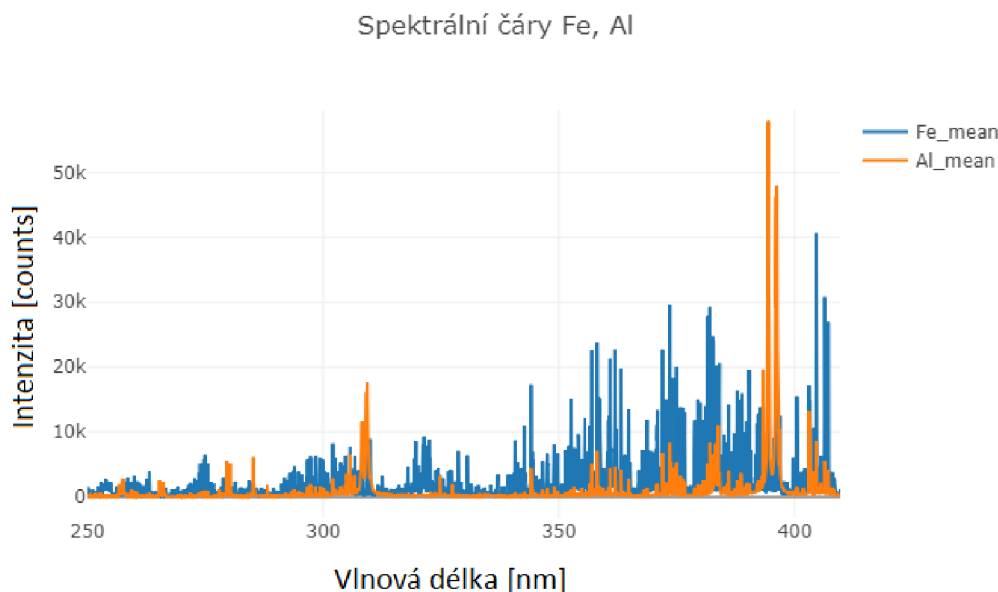
Tato metoda je velmi efektivní ve smyslu rychlého měření. Získáme velké množství dat, které budeme třídit a zpracovávat. Jednotlivé shluky u heterogenní matrice jsou propojeny, snažíme se nalézt hranici jednotlivých shluků.



Obrázek 23: Vizualizace dat v novém PC prostoru, a) vykreslení skóre první a čtvrté hlavní komponenty, b) materiálové přiřazení shluků na povrchu vzorku, c) charakteristika shluků pomocí prvků[36].

Mapujeme-li s vysokým rozlišením silně heterogenní vzorek, získáme velké množství dat. Pro heterogenní vzorek nám z 99,9% výstup z PCA neroztřídí jednotlivé prvky do každé dílčí komponenty separátně, ale pouze vyjádří vztah mezi jednotlivými významnými čarami, které ovšem charakterizují různé prvky. Například na obrázku 23 můžeme vidět silně heterogenní matrici horniny, která obsahuje pyrit, křemičitan a tyrkys [36]. Autor vykreslil první a čtvrtou hlavní komponentu. Vidíme, že se v datech nevytvořily opticky žádné shluky. Jde totiž o velmi heterogenní matrici. Za pomoci komponentních vah a vztahu jednotlivých komponent si autor subjektivně zvolil 4 shluky, tři z nich jsou charakteristické pro jistý materiál a čtvrtý (žlutý) shluk byl zvolen jako přechod těchto materiálů. Shluky následně vykreslil do 2D mapy povrchu vzorku (b) a v pravé části obrázku dole (c) charakterizoval jednotlivé shluky konkrétními materiály. Na obrázku 23 (a) vidíme, že nebyla roztržena všechna data, jaké bychom jim tedy měli přiřadit složení, na obrázku 23 (b) žádná hluchá místa nemáme.

V tomto případě bychom shluky nebyli schopni získat pomocí žádných algoritmů. V praktické části si zjednodušíme problém a budeme řešit pouze jeden vybraný přechod. Vybereme matrici dvou prvků u kterých máme jasně definované poměry. Zabýváme se matricí obsahující železo a hliník. Dva prvky, které mají velmi nevyrovnaná data jak do množství jednotlivých čar tak i jejich intenzit, viz obrázek 24. Provedeme diagnostiku, vizualizaci dat a následné rozdělení PCA prostoru podle jednotlivých prvků.



Obrázek 24: Vykreslení průměrných spektrálních čar Fe a Al. Vidíme zde rozdíl v počtu spektrálních čar a v intenzitách. Tento graf je pro přehlednost zúžený na interval vlnové délky od 250nm do 410nm .

4.1 Popis a detekce vzorku

Vzorkem použitým pro měření byl materiál skládající se ze dvou prvků, železa a hliníku. Tento vzorek měl jednoznačně určené rozhraní mezi jednotlivými prvky. Byl to vzorek válcovitého tvaru s průměrem cca 5cm (obrázek 27).

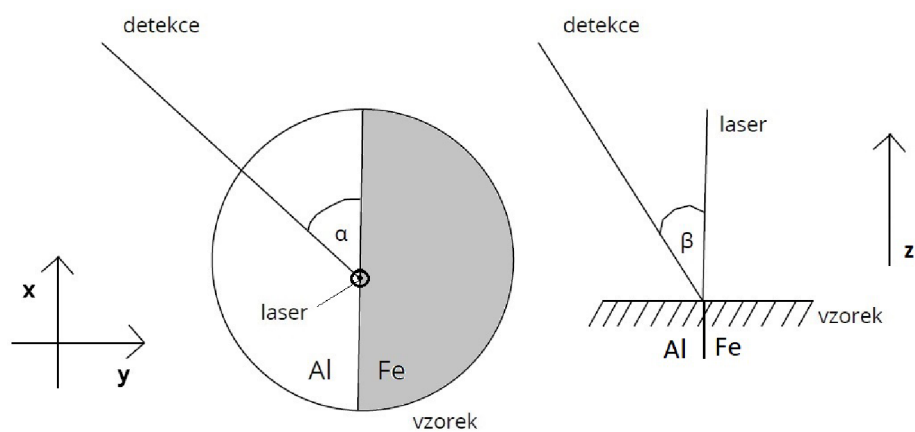
Soustředili jsme se na měření v oblasti rozhraní daných prvků v materiálu.

Laserový paprsek měl danou plochu pulzu, ze které se ablatoval materiál a vzniklo plazma. Tímto paprskem jsme postupně přecházeli hranici dvou různých materiálů s měnícím se plošným poměrem viz obrázek 26. Na tomto obrázku můžeme vidět, jak probíhalo měření. Naším úkolem bylo zaznamenat postupný přechod mezi dvěma materiály ve vzorku. Při každém dalším měření jsme proměňovali poměr prvků ve vzorku.

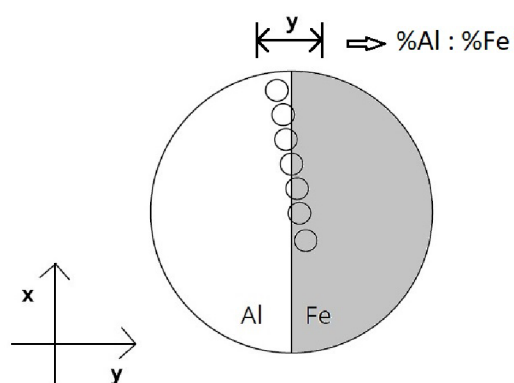
4.2 Přístroj

Na obrázku 27 můžeme vidět sestavu LIBS s detekcí na blízko. Tato sestava se nachází v laboratořích fyzikálního ústavu na FSI.

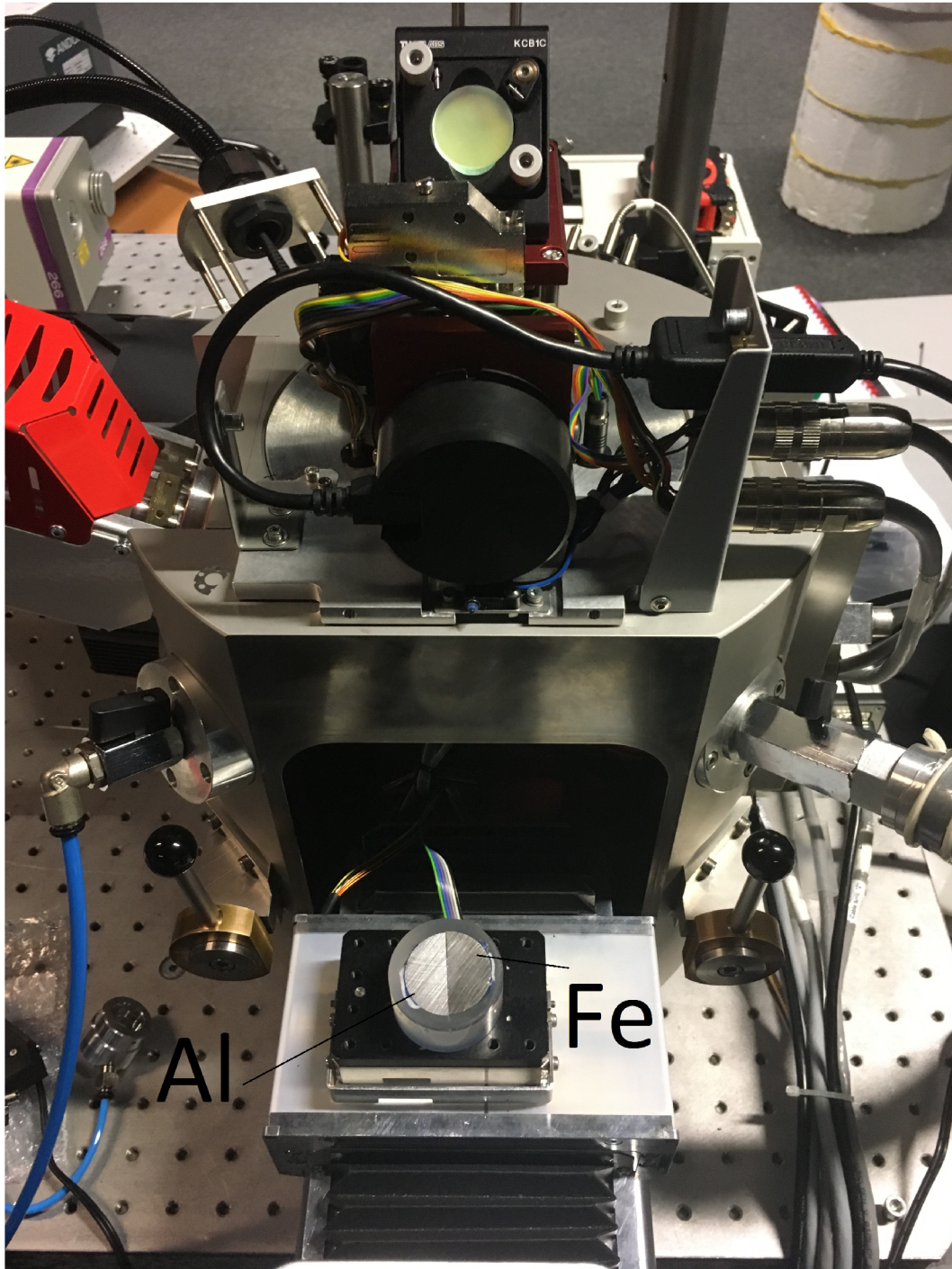
Při našem praktickém měření, které se uskutečnilo 13.3.2019 bylo stanoveno nastavení následovně: energie laserového pulsu měla hodnotu $E = 20\text{mJ}$, prodleva při snímání optické emise gate delay byla $1\mu\text{s}$, expoziční čas $50\mu\text{s}$, defokus $-0,5\text{mm}$ a laserový paprsek měl na povrchu vzorku průměr cca $250\mu\text{m}$.



Obrázek 25: Schématická vizualizace vzorku detektoru a laserového paprsku.



Obrázek 26: Schématické znázornění jednotlivých pulzů na vzorku.



Obrázek 27: Instrumentace aparatury včetně použitého vzorku.

4.3 Data

Výstupem z LIBS instrumentace dostaneme zdrojovou matici dat $X(n \times m)$ s n objekty a m znaky. Provedli jsme 1125 měření. Z těchto 1125 měření bylo provedeno každých 5 pulzů na totožném místě, poté byl paprsek laseru zaměřen na místo jiné. V každém poměru při měření hranice bylo naměřeno $25 \cdot 5 = 125$ měření na 25 místech. Pro každé měření detektor zaznamenal 65000 znaků (proměnných). Tyto proměnné jsou vlnové délky v rozsahu od $200,00nm$ do $1500,00nm$ s jemností měření $0,02nm$. Čisté prvky byly měřeny detailněji, $50 \cdot 5 = 250$ měření na 50 místech.

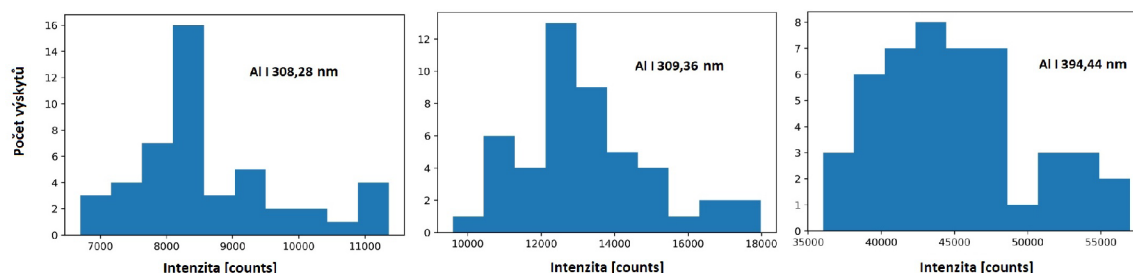
Máme tedy zdrojovou matici $X(1125 \times 65000)$. Jedná se o velmi rozsáhlou matici s velkým množstvím informací, se kterou budeme dále pracovat a získávat důležitá data pro interpretaci a vizualizaci.

4.4 Experiment

4.4.1 Test normality

První podstatnou informací je, zda náhodný výběr pochází z normálního rozdělení. Pro analýzu jsou důležitými spektrálními čarami takové, které odpovídají železným a hliníkovým čarám. Budeme tedy řešit, zda jsou data z normálního rozdělení pouze na těchto určitých čarách železa a hliníku.

Využijeme optický přístup ověření normality pomocí grafického znázornění distribuce jednotlivých čar.



Obrázek 28: Histogram distribuce rozdělení jednotlivých čar. Na ose x jsou intervalové rozsahy intenzit. Na ose y je počet výskytů v datovém souboru pro jednotlivé intervaly intenzit.

Posléze na data provedeme Shapiro-Wilkův test. Tento test jsme vybrali vzhledem k dobré vypovídající hodnotě pro malé výběry o velikosti $n \leq 50$.

vlnová délka λ [nm]	Al I 308,28	Al I 309,36	Al I 394,44
p-hodnota [-]	0,00824271	0,104225	0,14577
Nulová hypotéza H_0	zamítáme	nezamítáme	nezamítáme

$\alpha = 0,05$

Obrázek 29: Ukázka výstupu z Shapiro-Wilkova testu na významných čarách vzorku čistého Al.

Jako výstup z testu dostaneme p-hodnotu. Prahová hodnota pro stanovení hypotézy se volí $\alpha = 5\%$ (můžeme volit také 1% nebo 10%) a nazýváme ji hladinou významnosti daného testu.

Nulová hypotéza H_0 zní: vzorek má normální rozdělení na hladině významnosti α . Alternativní hypotézu volíme jako negaci, tedy H_A zní: vzorek nemá normální rozdělení na hladině významnosti α .

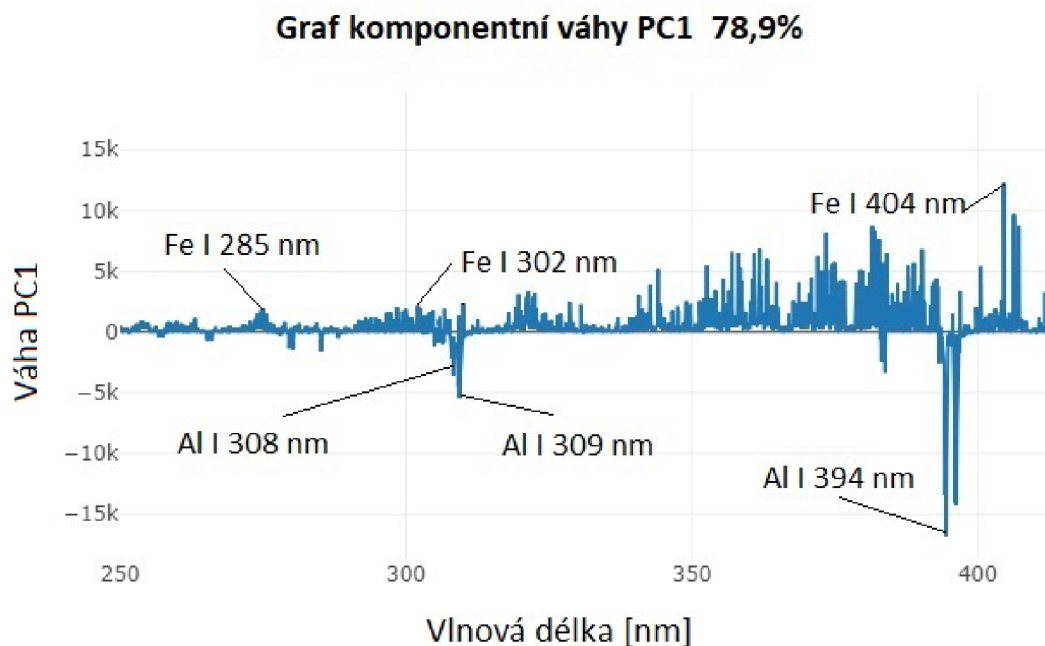
Pokud $p \leq \alpha$: zamítáme nulovou hypotézu H_0 a $p > \alpha$: nejsme schopni odmítnout H_0 . Nezamítnutí nulové hypotézy H_0 ještě neznamená, že hypotéza platí, neboť na základě realizace náhodného výběru jsme získali pouze informace, které nejsou dostatečné na zamítnutí.

Podle obrázku 29 se pohybujeme částečně na hranici normality, ale pro následné operace se zdrojovou maticí jsme nuceni brát předpoklad normálního rozdělení. Při nedodržení tohoto předpokladu nebudou data použitelná k dalšímu zpracování.

4.4.2 PCA

Na surová data aplikujeme metodu hlavních komponent. Tato metoda nám umožní první náhled do problematiky.

Jako první důležitý výstup z analýzy PCA jsou komponentní váhy.



Obrázek 30: Graf komponentní váhy pro první hlavní komponentu. Tento graf je pro přehlednost zúžený na interval vlnové délky od 250nm do 420nm.

Na obrázku 30 vidíme znázorněné podstatné čáry hliníku a železa. Můžeme si všimnout, že všechny čáry daného materiálu mají vždy orientaci ve směru osy y. Tedy při následné vizualizaci dat pomocí prvních dvou hlavních komponent můžeme předpokládat, že ve směru první hlavní komponenty se data rozmístí podle obsahu hliníku a železa ve vzorku. V záporné části první hlavní osy by se měla nacházet data, která obsahují více hliníku a v kladné části ty obsahující více železa.

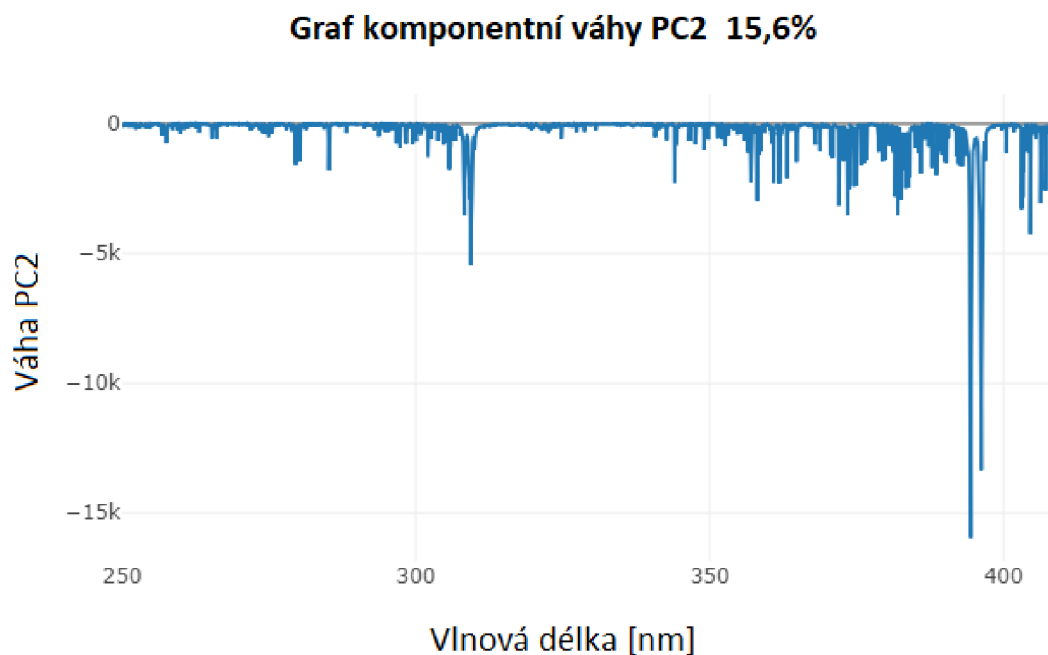
U druhé hlavní komponenty (obrázek 31) vidíme, že nám již nepřináší cenné informace pro analýzu. Všechny příspěvky vah jsou v negativním směru osy y. Takový průběh charakterizuje pouze fluktuaci v datech.

Data byla detekována na základě 5-ti pulzů v každém jednotlivém bodě. Prvními pulzy mohlo dojít k zaznamenání nečistot na povrchu vzorku a tím zkreslením údajů. Při vykreslení prvních pulzů jsme zaznamenali spektrální čáry prvků jako jsou Na, K, Ca, které se na pozdějších pulzech dále neobjevovaly.

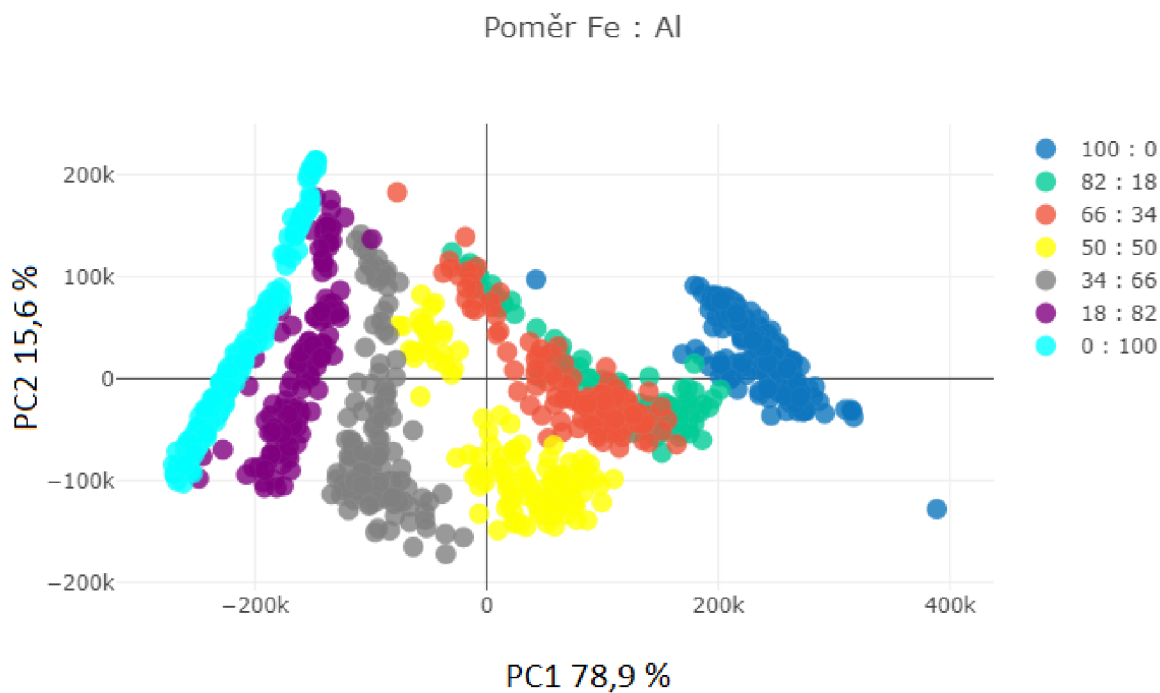
Dalším postupem našeho měření je selekce pulzu. Vybrán byl pulz třetí v pořadí, a na datech jím indukovaných provedeme analýzu hlavních komponent.

V tomto případě nám první hlavní komponenta vyšla podobně, ale s opačnou orientací vah pro hliník a železo, obrázek 33. Druhá hlavní komponenta vyšla jako fluktuace intenzity v datech.

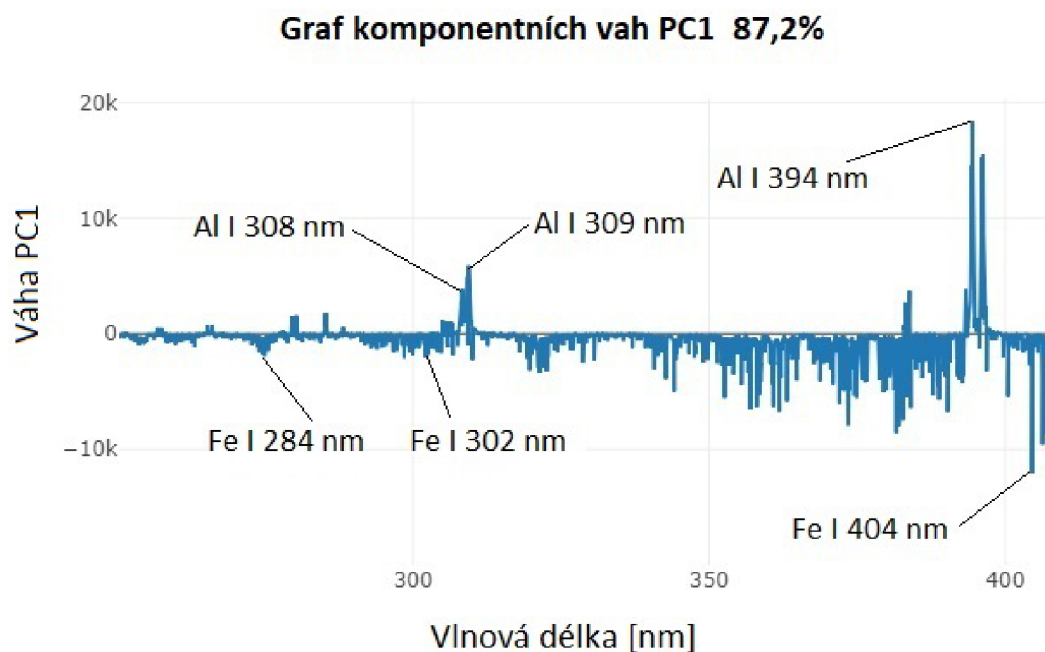
Na obrázku 35 vidíme lépe roztríděná data. Může to být způsobeno jak množstvím, tak



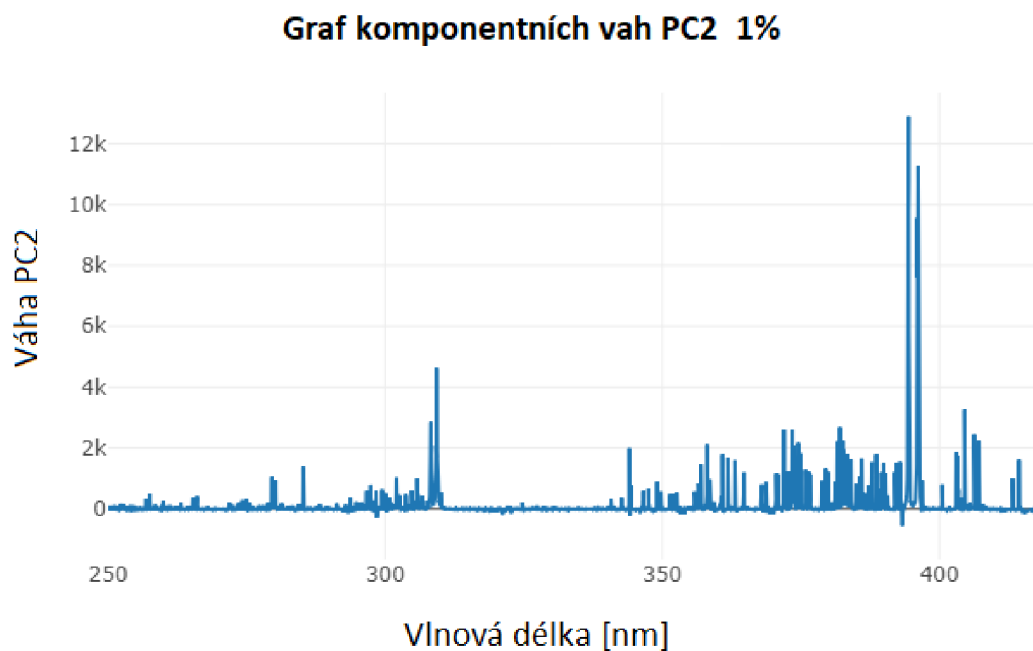
Obrázek 31: Graf komponentní váhy pro druhou hlavní komponentu. Tento graf je pro přehlednost zúžený na interval vlnové délky od 250nm do 420nm.



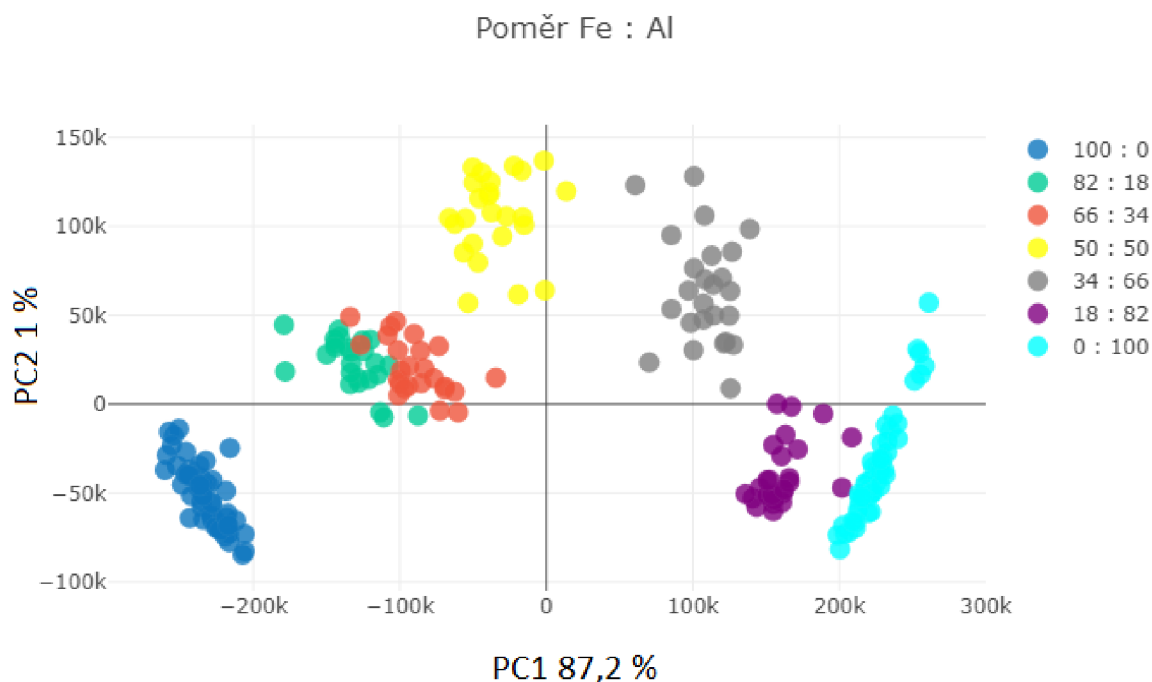
Obrázek 32: Vykreslení prvních dvou komponent PC1xPC2. Podle barvy jsou roztříděny jednotlivé poměry obsahu železa a hliníku.



Obrázek 33: Graf komponentní váhy pro první hlavní komponentu pro třetí pulz. Tento graf je pro přehlednost zúžený na interval vlnové délky od 250nm do 420nm.



Obrázek 34: Graf komponentní váhy pro druhou hlavní komponentu pro třetí pulz. Tento graf je pro přehlednost zúžený na interval vlnové délky od 250nm do 420nm.



Obrázek 35: Vykreslení prvních dvou komponent PC1xPC2 pro třetí pulz. Podle barvy jsou rozříděny jednotlivé poměry obsahu železa a hliníku.

i našim výběrem konkrétního pulzu.

Nastíháme si problém, kdy máme dva materiály železo a hliník, a daný vzorek měření, který je směsí těchto dvou prvků. Budeme postupovat analogicky a ukážeme si to například na směsi, která má poměr obou materiálů stejný. Provedeme analýzu PCA na třetím pulzu měření, jak bylo výše zmíněno. Komponentní váhy jsou podobné jako výše, tedy v kladném směru první hlavní komponenty je významný hliník a v záporném železo.

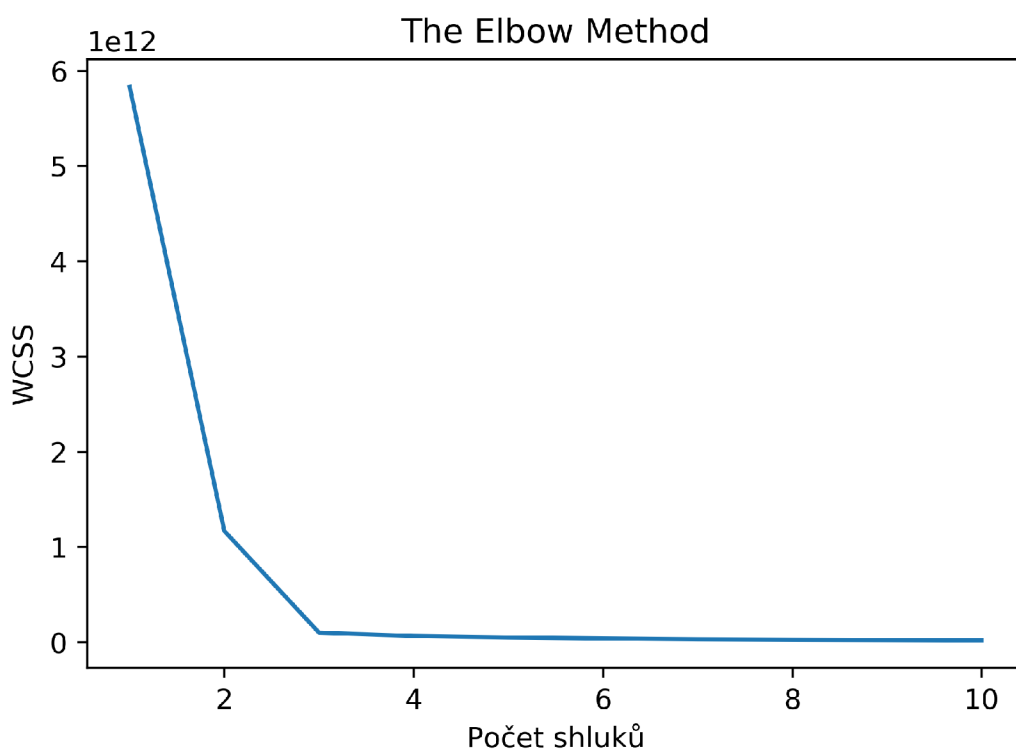
Dostaneme vizualizaci prvních dvou hlavních komponent, u kterých využijeme další analytickou metodu clusterů. První otázkou je, jaký je optimální počet clusterů? Elbow metoda nám s tím pomůže. Tato metoda využívá chybu součtů uvnitř clusteru (WCSS - with cluster sum of errors). WCSS je součet vzdáleností bodů shluku od centrálního bodu daného shluku.

Pokud WCSS vykreslíme ve vztahu s počtem shluků, dostaneme graf známý jako The Elbow Method. V grafu pozorujeme, při jakém počtu shluků vzniká zlom.

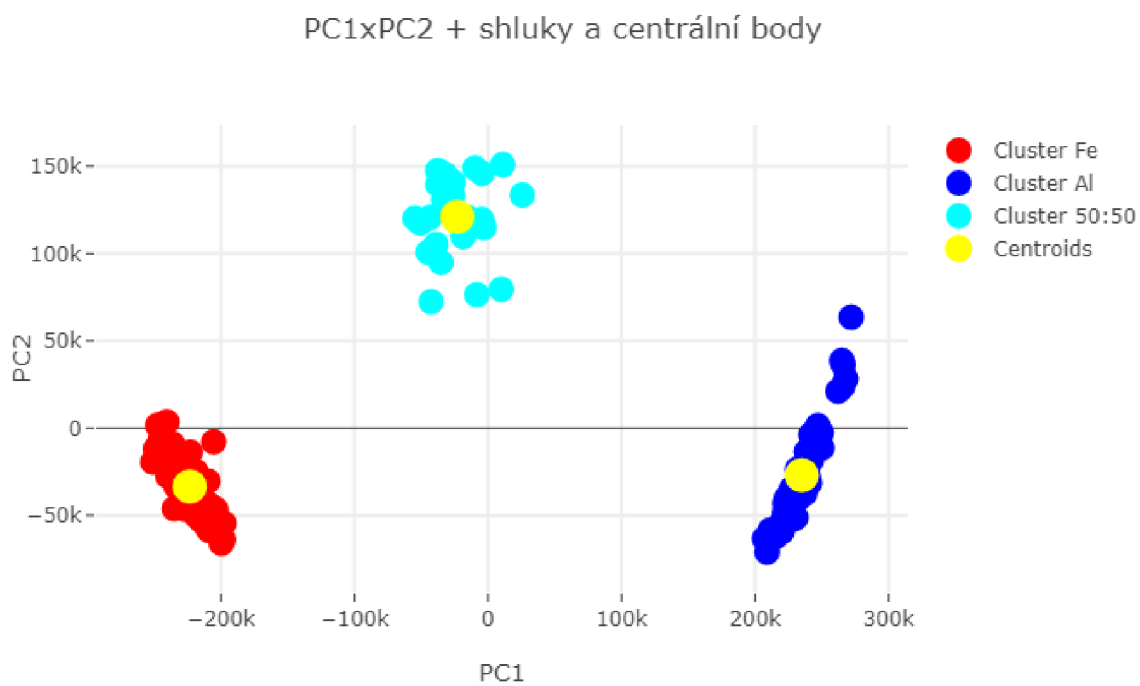
Z obrázku 37 jsme dostali jasný výstup pro zvolení optimálního počtu shluků na tři. Pomocí k-means clusteru napočítáme tři shluky a jejich centrální body, které vykreslíme do grafu prvních dvou hlavních komponent.

Následuje otázka, jak cluster směsi ohodnotit a jakou mu přiřadit hodnotu poměru složení jednotlivých materiálů. Tento problém budeme řešit podle vzdálenosti centrálního bodu směsi od centrálních bodů čistých prvků. Využijeme Euklidovskou metriku a vytvoříme bar chart, ve kterém budou jednotlivé vzdálenosti zobrazeny.

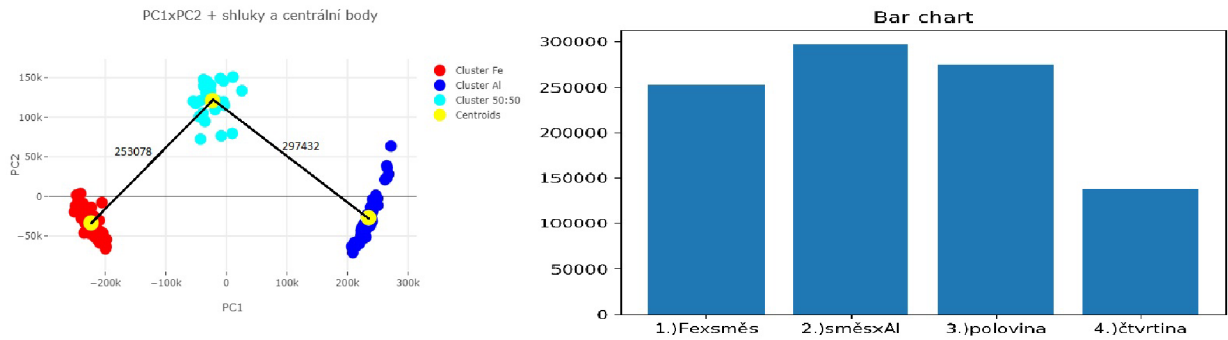
Obrázek 38 nám dává představu o materiálovém složení vzorku. Vzdálenost udává z čeho se směs skládá. Materiál obsahuje více prvek, ke kterému je blíže vyobrazen. Pomocný bar chart ukazuje další dvě hodnoty. Polovina která udává přesný poměr 50:50 a čtvrtina která udává poměr 75:25.



Obrázek 36: Graf The Elbow Method pro určení počtu shluků. Vidíme jasný lom u počtu 3.

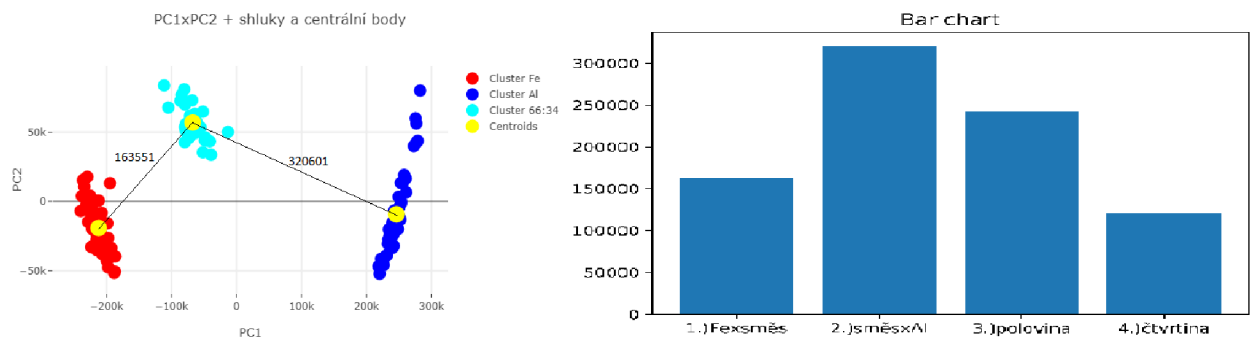


Obrázek 37: Vizualizace dat s rozříděním do clusterů a vytvořením centrálních bodů clusterů.

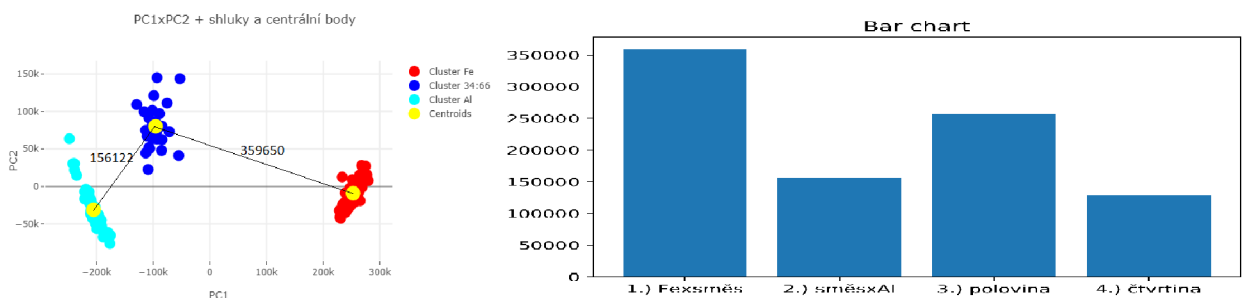


Obrázek 38: Euklidovská vzdálenost centrálních bodů.

Zkoumaná směs byla v poměru 50:50, tedy v ideálním případě se měl shluk směsi zobrazit ve stejné vzdálenosti od zbylých dvou shluků. Nicméně jsme získali poměrně zajímavý výstup o neznámé směsi, kterým bychom i pomocí bar chartu hodnotili poměr kolem poloviny, mírně se klonící k většímu obsahu železa ve směsi. Můžeme si ukázat i další směsi a jejich výslednou analýzu.



Obrázek 39: Euklidovská vzdálenost centrálních bodů pro směs 66:34.



Obrázek 40: Euklidovská vzdálenost centrálních bodů pro směs 34:66.

Tabulka 2: V prvním řádku jsou teoretické poměry získané měřením ablace na rozhraní a ve druhém jsou výsledky, které jsme dostali při zpracování.

teoretický poměr Fe : Al	82 : 18	66 : 34	50 : 50	34 : 66	18 : 82
poměr získaný analýzou Fe : Al	74,0 : 26,0	66,2 : 33,8	54,0 : 46,0	30,3 : 69,7	14,8 : 85,2

V tabulce 2 vidíme jednotlivé poměry směsí a výsledky po analýze. Tuto tabulku jsme získali kompletací předchozích postupů na různé směsi. Teoretické výsledky v tabulce nejsou exaktně přesné. K nepřesnosti dochází při nastavování laserového paprsku (obrázek: 26) a také při záznamu emisního spektra detektorem (obrázek: 25). Tento detektor zaznamenává data pod úhlem vychýleným na jednu stranu. Pokud tedy není ablatované plazma dokonale homogenní (promísené), může docházet k chybě při sbírání dat. K dalšímu zkreslení také dochází z fyzikálně-chemického pohledu, díky velké diverzitě prvků železa a hliníku.

Analýza hlavních komponent přinesla informačně významné údaje. Lepší výsledky dostaneme při podobných poměrech dvou prvků (50:50, 66:34). Body jsou rozmístěny v prostoru tak, že nemáme problém s kvalifikací těchto bodů do jednotlivých shluků. Pokud máme směs která má větší poměr jednoho ze dvou vzorku (18:82), dostáváme body blízko sebe, které se nám hůře separují do shluků. Tímto je způsobena menší přesnost analýzy u směsí, ve kterých výrazně převažuje jeden prvek.

4.4.3 Korelační analýza

K získání vztahů mezi jednotlivými čarami využijeme korelační analýzu.

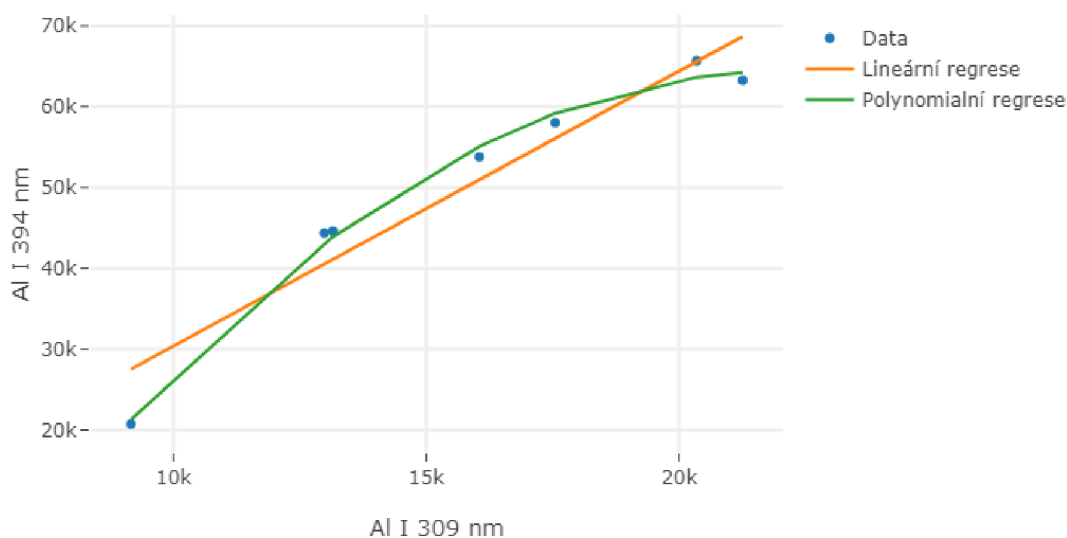
Vytvoříme korelační matici pro významné čáry hliníku a železa. Hliník a železo mají velmi nevyrovnaná data co do počtu charakteristických spektrálních čar. Z tohoto důvodu nebylo možné u hliníku najít více vhodných spektrálních čar. Na obrázku 41 vidíme velkou korelaci (závislost) jednotlivých čar hliníku mezi sebou a jednotlivých čar železa mezi sebou.

Index		Al			Fe				
		308.28 nm	309.36 nm	394.44 nm	253.72 nm	371.62 nm	538.3 nm	492.02 nm	489.14 nm
Al	308.28 nm	1	0.995	0.984	-0.833	-0.859	-0.886	-0.835	-0.847
	309.36 nm	0.995	1	0.985	-0.825	-0.852	-0.88	-0.827	-0.84
	394.44 nm	0.984	0.985	1	-0.789	-0.815	-0.848	-0.777	-0.793
Fe	253.72 nm	-0.833	-0.825	-0.789	1	0.962	0.961	0.964	0.966
	371.62 nm	-0.859	-0.852	-0.815	0.962	1	0.985	0.987	0.989
	538.3 nm	-0.886	-0.88	-0.848	0.961	0.985	1	0.984	0.988
	492.02 nm	-0.835	-0.827	-0.777	0.964	0.987	0.984	1	0.998
	489.14 nm	-0.847	-0.84	-0.793	0.966	0.989	0.988	0.998	1

Obrázek 41: Korelační matice.

Dále si můžeme rozebrat vzájemný vztah hliníkových čar Al I 309 nm, Al I 394 nm při zvyšování poměru železa ve vzorku. Na obrázku 42 vidíme, že při nízkém poměru hliníku ve vzorku je závislost téměř lineární. Tato vlastnost se se zvyšujícím poměrem postupně ztrácí a přechází spíše v závislost na vyšším řádu. To je způsobeno saturací čáry Al I 394 nm při vyšším obsahu hliníku.

Intenzity spektrálních čar hliníku při zvyšování poměru Al ve vzorku



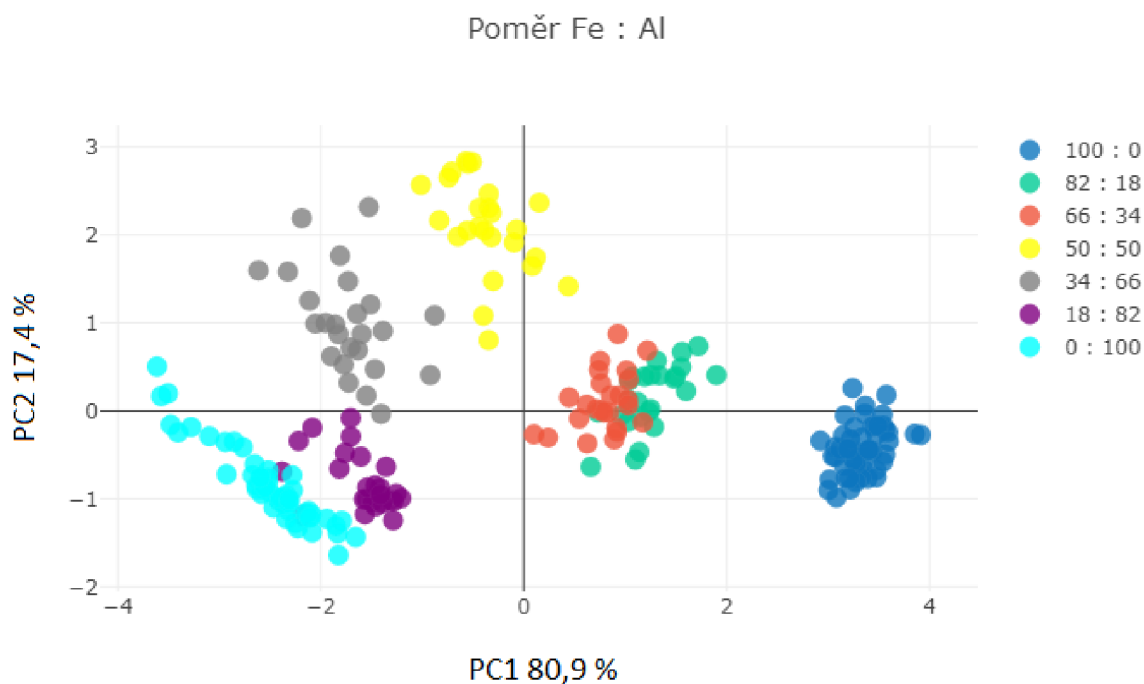
Obrázek 42: Graf závislosti Al I 309 a Al I 394. Vytvořena lineární regrese a polynomiální regrese druhého řádu.

4.4.4 Alternativní přístup

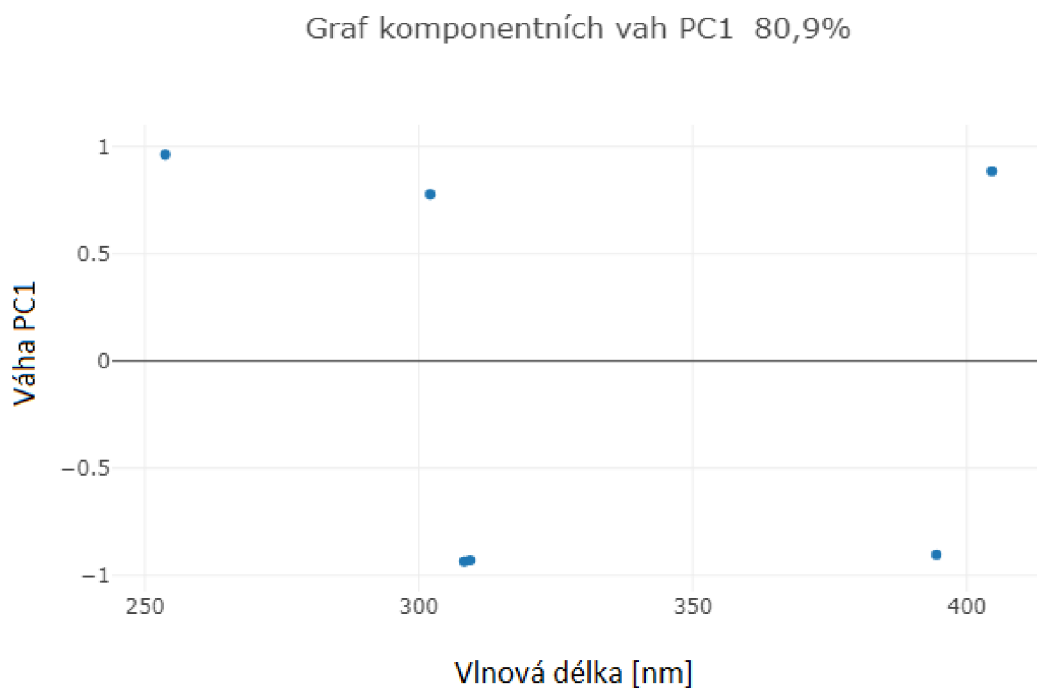
Alternativním přístupem k analýze je výběr významných spektrálních čar jednotlivých prvků. Například těch získaných korelační analýzou. Tento výběr snížil dimenzi úlohy již na začátku. Došlo ke zjednodušení výpočtové náročnosti v závislosti ztráty informací z dat. Výběr nám také odstranil šum v datech, která nyní můžeme normalizovat. Pokračujeme totožným postupem, vykreslíme první dvě hlavní komponenty pro různé směsi a k nim odpovídající komponentní váhy.

Graf první hlavní komponenty nám ukazuje, že v kladném směru PC1 bude roztríděno železo a v záporném směru hliník. Druhá hlavní komponenta nám vyjádřila fluktuaci v datech.

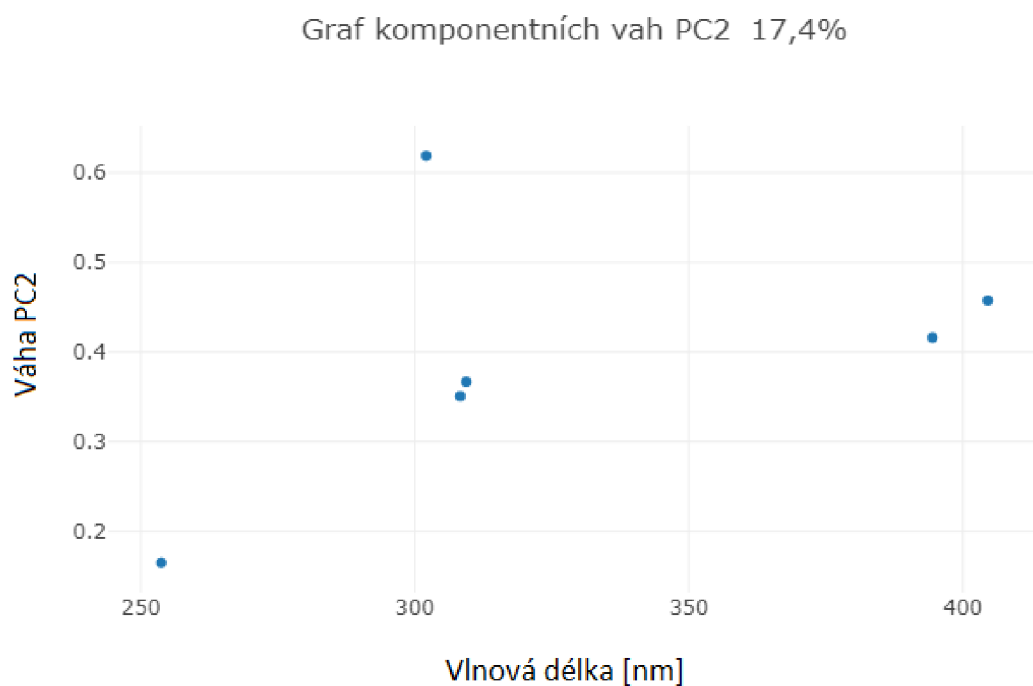
Vybereme směs se stejným poměrem železa a hliníku, vytvoříme shluky a najdeme centrální body. Na obrázku 46 vidíme, že jsou si výstupy velmi podobné jako při postupu s velkou dimenzí výše. Snížení dimenze se projevilo negativně, zároveň ale mohlo dojít k normalizaci z důvodu odstranění šumu v datech. Více různých kombinací zpracování viz tabulka 4 dále.



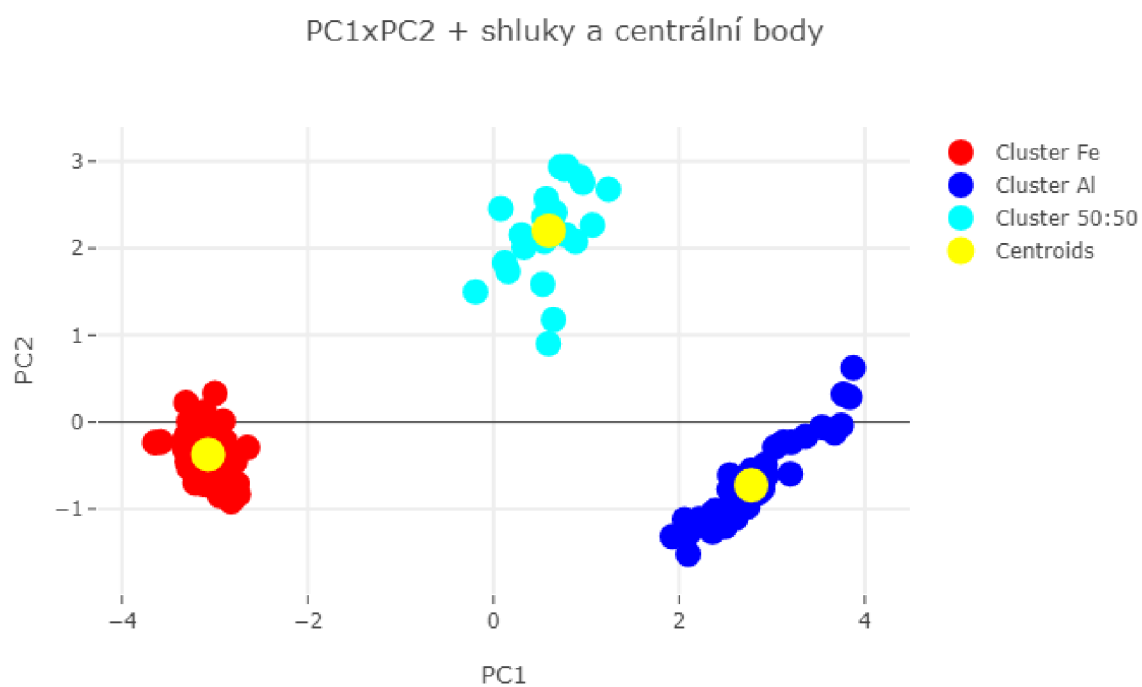
Obrázek 43: Vykreslení prvních dvou komponent PC1xPC2 pro třetí pulz se sníženou dimenzí vstupních dat. Podle barvy jsou roztrženy jednotlivé poměry obsahu železa a hliníku.



Obrázek 44: Graf komponentní váhy pro první hlavní komponentu.



Obrázek 45: Graf komponentní váhy pro druhou hlavní komponentu.



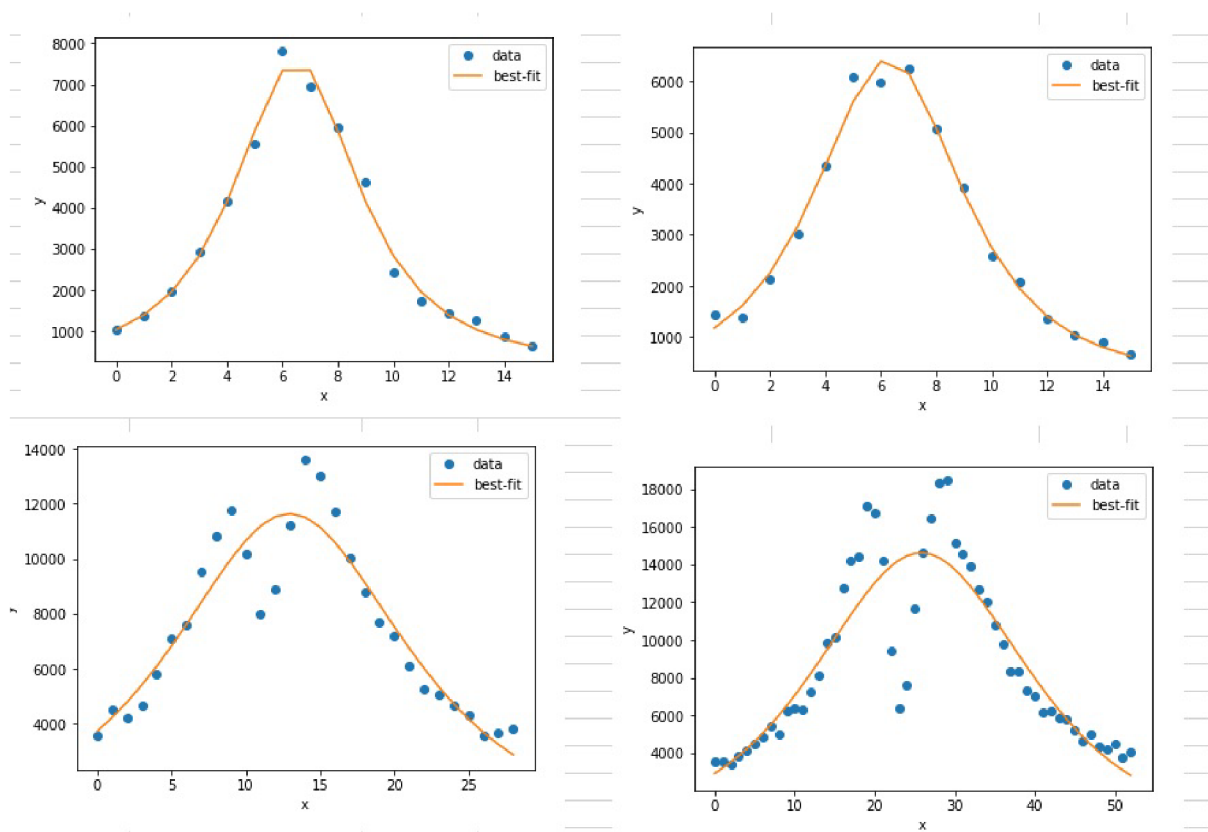
Obrázek 46: Vykreslení prvních dvou komponent PC1xPC2 pro třetí pulz se sníženou dimenzí. Vytvoření třech shluků a jejich centrálních bodů.

Tabulka 3: Tabulka kombinací vstupních dat a typů zpracování s výstupem obsahujícím stejný poměr železa a hliníku.

počet proměnných	typ zpracování	výběr čar	obsah železa ve směsi	obsah hliníku ve směsi
65000	-	-	0.540	0.460
	centrování	-	0.540	0.460
	centrování + outlier	-	0.499	0.501
	normalizace	-	0.638	0.362
3Fe + 3Al	-	Max	0.330	0.670
	-	Voigt	0.197	0.803
	normalizace	Max	0.449	0.551
5Fe + 3Al	-	Max	0.364	0.636
	normalizace	Max	0.460	0.540
12Fe + 3Al	-	Max	0.465	0.535
	normalizace	Max	0.566	0.434
1Fe + 1Al	-	Max	0.614	0.386
	normalizace	Max	0.474	0.526

Různé kombinace předzpracování a výběru dat již nebudeme více detailněji popisovat, ale vytvoříme tabulku s kombinacemi těchto operací a výstupy pro směs obsahující 50% železa a 50% hliníku.

V této tabulce 3 pojem normalizace myslíme standardizací směrodatnou odchylkou. Další typy normalizace, které se dají použít je, že původní hodnotu vydělíme celkovou intenzitou spektra nebo nějakou referenční čarou, v našem případě by šlo o čáru železa nebo hliníku. V tabulce jsme prováděli aproximace Voigtovou funkcí, dále také vhodné vyzkoušet aproximace Lorenzovou či Gaussovou funkcí, tyto typy aproximace nám neposkytly uspokojivý výsledek, kvůli významné stauraci spektrálních čar našeho vzorku (obrázek 47).



Obrázek 47: Ukázka saturace čar při aproximaci Voigtovou funkcí.

Tabulka 4: Výběr spektrálních čar vhodných pro analýzu. Tyto čáry byly použity v tabulce 3.

prvek	stupeň ionizace	vlnová délka [nm]	$A_{ki} [10^8 \cdot s^{-1}]$	E_i [eV]	E_k [eV]
Al	I	308.21	0.59	0	4.0215
	I	309.27	0.73	0.0138	4.0217
	I	394.40	0.50	0.0000	3.1427
Fe	I	253.72	3.7	2.4040	7.2890
	I	284.36	1.12	1.4850	4.7330
	I	302.06	0.76	0.0000	4.1030
	I	344.06	0.17	0.0000	3.6025
	I	347.55	0.10	0.0873	3.6537
	I	371.64	0.35	2.9398	6.2750
	I	404.58	0.86	1.4849	4.5485
	I	427.16	0.23	1.4849	4.5485
	I	440.48	0.28	1.5574	4.3714
	I	489.14	0.31	2.8512	5.3852
	I	492.02	0.36	2.8325	5.3516
	I	538.30	0.78	4.3125	6.6149

5 Závěr

Spektroskopie laserem buzeného plazmatu při použití vhodných chemometrických metod se ukázala jako vhodný nástroj k určování prvkového složení vzorků. Jejimi pozitivy je rychlost získání dat a přínos velmi rychlých výsledků.

Praktická část práce byla zaměřena na analýzu heterogenního materiálu na rozhraní matic dvou prvků (hliníku a železa). Při analýze tohoto přechodu docházelo k propojení shluků, které bylo nutné vhodně klasifikovat.

Po provedení analýzy hlavních komponent jsme získali informaci o rozložení jednotlivých měření v novém PCA prostoru. Zjistili jsme, že není vhodné zabývat se kompletní zdrojovou maticí dat, ale pouze jedním dílčím pulzem při daném měření. Pro takto námi vybraná data jsme následně provedli kompletní analýzu s vytvořením shluků a materiálovou kvalifikaci popř. kvantifikaci každého shluku viz tabulka 2. V této tabulce jsme dostali informačně významné údaje o obsazích jednotlivých materiálů.

Dále byla provedena korelační analýza, která nám potvrdila předpoklad závislosti jednotlivých čar pro určitý prvek. Díky této informaci jsme mohli podstatně zredukovat dimenzi úlohy hned na jejím začátku. Volili jsme různé počty hliníkových a železných spektrálních čar, na kterých jsme realizovali analýzu. Vytvořili jsme tabulku 3 ve které jsme prováděli různé operace s různou zdrojovou maticí dat a analyzovali jednotlivé přístupy na směsi se stejným poměrem obou prvků. Data v této tabulce dedukují konkluze této práce. Centrování dat nemá vliv na výsledek, nicméně nám ukáže rozdíly v intenzitách jednotlivých spektrálních čar. Odstranění outlierů zlepšilo výsledek analýzy. Normalizace na data, která nejsou zatížena šumem se při analýze projeví také pozitivně. Při výběru konkrétního počtu proměnných se s množstvím čar mírně zlepšila analýza. Toto tvrzení však neplatí obecně. Každý prvek je určen významnějšími a méně významnými čarami, kde ale nastává problém jednoznačného určení významnosti konkrétní čáry. Při provádění aproximace zdrojových dat pomocí Lorenzovy, Gaussovy či Voigtovy funkce jsme nedosáhli předpokládaných výsledků z důvodu nevhodnosti použití těchto přístupů na naše konkrétní data, která byla významně saturována. Prováděli jsme tedy aproximaci výběrem maxima z intervalu oblasti čáry. I přes ztrátu dat pomocí této aproximace jsme dosáhli zajímavých výsledků.

6 Zdroje

Reference

- [1] MELOUN, Milan, Jiří MILITKÝ a Martin HILL: *Statistická analýza vícerozměrných dat v příkladech*. V Praze: Univerzita Karlova, nakladatelství Karolinum, 2017. ISBN 978-80-246-3618-4.
- [2] LEVER, Jake, Martin KRZYWINSKI a Naomi ALTMAN: *Principal component analysis*. Nature Methods [online]. 2017, 14(7), 641-642 [cit. 2019-05-07]. DOI: 10.1038/nmeth.4346. ISSN 1548-7091. Dostupné z: <http://www.nature.com/articles/nmeth.4346>.
- [3] ABDI, Hervé a Lynne J. WILLIAMS: *Principal component analysis*. Wiley Interdisciplinary Reviews: Computational Statistics [online]. 2010, 2(4), 433-459 [cit. 2019-05-07]. DOI: 10.1002/wics.101. ISSN 19395108. Dostupné z: <http://doi.wiley.com/10.1002/wics.101>.
- [4] MELOUN, Milan a Jiří MILITKÝ: *Statistická analýza experimentálních dat*. Vyd. 2., upr. a rozš. Praha: Academia, 2004. ISBN 80200-1254-0.
- [5] BRO, Rasmus a Age K. SMILDE: *Principal component analysis*. Anal. Methods [online]. 2014, 6(9), 2812-2831 [cit. 2019-05-02]. DOI: 10.1039/C3AY41907J. ISSN 1759-9660. Dostupné z: <http://pubs.rsc.org/-/content/articlehtml/2014/ay/c3ay41907j>, vol. 6, str. 2812-2831.
- [6] POŘÍZKA, Pavel, Jakub KLUS, Erik KÉPEŠ, David PROCHAZKA, David W. HAHN a Jozef KAISER: *On the utilization of principal component analysis in laser-induced breakdown spectroscopy data analysis, a review*. Spectrochimica Acta Part B: Atomic Spectroscopy [online]. 2018, 148, 65-82 [cit. 2019-05-07]. DOI: 10.1016/j.sab.2018.05.030. ISSN 05848547. Dostupné z: <https://linkinghub.elsevier.com/retrieve/pii/S0584854718301526>.
- [7] POŘÍZKA, P., J. KLUS, A. HRDLIČKA, et al: *Impact of Laser-Induced Breakdown Spectroscopy data normalization on multivariate classification accuracy*. Journal of Analytical Atomic Spectrometry [online]. 2017, 32(2), 277-288 [cit. 2019-05-06]. DOI: 10.1039/C6JA00322B. ISSN 0267-9477.
- [8] NOLL, Reinhard: *Laser-Induced Breakdown Spectroscopy* [online]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012 [cit. 2019-05-07]. DOI: 10.1007/978-3-642-20668-9. ISBN 978-3-642-20667-2.
- [9] WINEFORDNER, James D., Igor B. GORNUSHKIN, Tiffany CORRELL, Emily GIBB, Benjamin W. SMITH a Nicol OMENETTO: *Comparing several atomic spectrometric methods to the super stars: special emphasis on laser induced breakdown spectroscopy, LIBS, a future super star*. Journal of Analytical Atomic Spectrometry. 2004, 19(9). DOI: 10.1039/b400355c. ISSN 0267-9477.

- [10] HALLIDAY, David, Robert RESNICK a Jearl WALKER, DUB, Petr, ed: *Fyzika*. 2., přeprac. vyd. Brno: VUTIUM, c2013. Překlady vysokoškolských učebnic. ISBN 978-80-214-4123-1.
- [11] *Spectral-Lines Analysis [online]*. China: Lifeng, 2007 [cit. 2019-05-02]. Dostupné z: <https://lifeng.lamost.org/courses/astrotoday/CHAISSON/AT304/HTML/AT30404.HTM>.
- [12] HRDLIČKOVÁ, Zuzana: *Základní rozdělení pravděpodobnosti. Matematika online [online]*. Brno: Ústav Matematiky FSI Brno, 2005 [cit. 2019-05-02]. Dostupné z: <http://mathonline.fme.vutbr.cz/Zakladni-rozdeleni-pravdepodobnosti/sc-1206-sr-1-a-200/default.aspx>.
- [13] RAUTIAN, Sergei G a Igor SOBEL'MAN: *THE EFFECT OF COLLISIONS ON THE DOPPLER BROADENING OF SPECTRAL LINES*. 1967 Sov. Phys. Usp. 9 701. Soviet Physics Uspekhi [online]. 1967, 9(5), 701-716 [cit. 2019-05-07]. DOI: 10.1070/PU1967v009n05ABEH003212. ISSN 0038-5670. Dostupné z: <http://stacks.iop.org/0038-5670/9/i=5/a=R04?key=crossref.9451705b8f7063b47912538c46846b0c>.
- [14] TORRES, J, J JONKERS, M J van de SANDE, J J A M van der MULLEN, A GAMERO a A SOLA: *An easy way to determine simultaneously the electron density and temperature in high-pressure plasmas by using Stark broadening*. Journal of Physics D: Applied Physics [online]. 2003, 36(13), L55-L59 [cit. 2019-05-07]. DOI: 10.1088/0022-3727/36/13/101. ISSN 0022-3727. Dostupné z: <http://stacks.iop.org/0022-3727/36/i=13/a=101?key=crossref.5ac9d3f8624ace9932e8af2f712e40ed>
- [15] GRIEM, Hans R: *Spectral line broadening by plasmas*. New York: Academic Press, 1974. ISBN 0123028507. Academic, New York, 1974.
- [16] Weisstein, Eric W: *Convolution*, From MathWorld—A Wolfram Web Resource. Dostupné z: <http://mathworld.wolfram.com/Convolution.html>.
- [17] *Stimulovaná emise*. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001-2019 [cit. 2019-05-02]. Dostupné z: https://cs.wikipedia.org/wiki/Stimulovan%C3%A1_emise.
- [18] *Diracovo delta*. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001-2019 [cit. 2019-05-02]. Dostupné z: https://cs.wikipedia.org/wiki/Diracovo_delta.
- [19] UDELL, Madeleine, Corinne HORN, Reza ZADEH a Stephen BOYD: *Generalized Low Rank Models*. Foundations and Trends® in Machine Learning [online]. 2016, 9(1), 1-118 [cit. 2019-05-07]. DOI: 10.1561/22000000055. ISSN 1935-8237. Dostupné z: <http://www.nowpublishers.com/article/Details/MAL-055>.
- [20] MIZIOLEK, Andrzej W., V. PALLESECHI a Israel SCHECHTER: *Laser-induced breakdown spectroscopy (LIBS): fundamentals and applications*. New York: Cambridge University Press, 2006. ISBN 9780521852746.

- [21] SATTMANN, R, V STURM a R NOLL: *Laser-induced breakdown spectroscopy of steel samples using multiple Q-switch Nd: YAG laser pulses*. Journal of Physics D: Applied Physics [online]. 1995, 28(10), 2181-2187 [cit. 2019-05-06]. DOI: 10.1088/0022-3727/28/10/030. ISSN 0022-3727. Dostupné z: <http://stacks.iop.org/0022-3727/28/i=10/a=030?key=crossref.0acf5f7e53c3a34b438c53554e214d92>
- [22] STEHRER, T., B. PRAHER, R. VISKUP, J. JASIK, H. WOLFMEIR, E. ARENHOLZ, J. HEITZ a J. D. PEDARNIG: *Laser-induced breakdown spectroscopy of iron oxide powder*. Journal of Analytical Atomic Spectrometry [online]. 2009, 24(7) [cit. 2019-05-06]. DOI: 10.1039/b817279j. ISSN 0267-9477. Dostupné z: <http://xlink.rsc.org/?DOI=b817279j>
- [23] KAISER, Jozef: *Spektroskopie laserem buzeného plazmatu: Laser-induced breakdown spectroscopy : teze přednášky k profesorskému jmenovacímu řízení v oboru aplikovaná fyzika*. Brno: VUTIUM, 2012. ISBN 978-80-214-4423-2.
- [24] BARBINI, R, F COLAO, V LAZIC, R FANTONI, A PALUCCI a M ANGELONE: *On board LIBS analysis of marine sediments collected during the XVI Italian campaign in Antarctica*. Spectrochimica Acta Part B: Atomic Spectroscopy [online]. 2002, 57(7), 1203-1218 [cit. 2019-05-07]. DOI: 10.1016/S0584-8547(02)00055-1. ISSN 05848547. Dostupné z: <https://linkinghub.elsevier.com/retrieve/pii/S0584854702000551>
- [25] RADZIEMSKI, Leon J: *From LASER to LIBS, the path of technology development*. Spectrochimica Acta Part B: Atomic Spectroscopy [online]. 2002, 57(7), 1109-1113 [cit. 2019-05-07]. DOI: 10.1016/S0584-8547(02)00052-6. ISSN 05848547. Dostupné z: <https://linkinghub.elsevier.com/retrieve/pii/S0584854702000526>
- [26] VÍTKOVÁ, Gabriela: *Spektrometrie laserem buzeného plazmatu (LIBS) vzdálených objektů [online]*. Brno, 2015 [cit. 2019-05-06]. Dostupné z: <https://is.muni.cz/th/b10me/>. Disertační práce. Masarykova univerzita, Přírodovědecká fakulta. Vedoucí práce Karel Novotný.
- [27] GHOLAP, Deepti S., Andrei IZMER, Björn DE SAMBER, et al.: *Comparison of laser ablation-inductively coupled plasma-mass spectrometry and micro-X-ray fluorescence spectrometry for elemental imaging in *Daphnia magna**. Analytica Chimica Acta [online]. 2010, 664(1), 19-26 [cit. 2019-05-21]. DOI: 10.1016/j.aca.2010.01.052. ISSN 00032670. Dostupné z: <https://linkinghub.elsevier.com/retrieve/pii/S0003267010001546>.
- [28] R. J. Goldston and Paul H Rutherford: *INTRODUCTION TO PLASMA PHYSICS*. Institute of Physics Publishing Bristol and Philadelphia, IOP Publishing Ltd 1995, ISBN 0 7503 0325 5.
- [29] NAMIHIRA, Takao, Shunsuke SAKAI, Takahiro YAMAGUCHI, et al.: *Electron Temperature and Electron Density of Underwater Pulsed Discharge Plasma Produced by Solid-State Pulsed-Power Generator*. IEEE Transactions on Plasma Science [online]. 2007, 35(3), 614-618 [cit.

2019-05-07]. DOI: 10.1109/TPS.2007.896965. ISSN 0093-3813. Dostupné z: <http://ieeexplore.ieee.org/document/4237270/>.

- [30] V. Unnikrishnan , K. Alti , V. Kartha, C. Santhosh, G. Gupta and B. Suri: *Measurements of plasma temperature and electron density in laser-induced copper plasma by time-resolved spectroscopy of neutral atom and ion emissions*. Institute of Physics Publishing Bristol and Philadelphia, IOP Publishing Ltd 1995, ISBN 0 7503 0325 5.
- [31] HAN, Jiawei a Micheline KAMBER: *Data mining: concepts and techniques*. 3rd ed. Burlington, MA: Elsevier, c2012. ISBN 978-0-12-381479-1.
- [32] O'Reilly Media: *Big Data Now: 2016 Edition*. O'Reilly Media, Inc.,1005 Gravenstein Highway North, Sebastopol, CA 95472., 2017. ISBN: 9781492049197
- [33] KIRK, Matthew: *Thoughtful machine learning*. Beijing: O'Reilly, [2015]. ISBN 9781449374068.
- [34] *Interquartile range*. In: Wikipedia: the free encyclopedia [online]. San Francisco (CA): Wikimedia Foundation, 2001-2019 [cit. 2019-05-02]. Dostupné z: https://en.wikipedia.org/wiki/Interquartile_range.
- [35] KÉPEŠ, Erik, Pavel POŘÍZKA, Jakub KLUS, Pavlína MODLITBOVÁ a Jozef KAISER: *Influence of baseline subtraction on laser-induced breakdown spectroscopic data*. Journal of Analytical Atomic Spectrometry [online]. 2018, 33(12), 2107-2115 [cit. 2019-05-07]. DOI: 10.1039/C8JA00209F. ISSN 0267-9477. Dostupné z:<http://xlink.rsc.org/?DOI=C8JA00209F>.
- [36] MONCAYO, Samuel, Ludovic DUPONCHEL, Niloofar MOUSAVIPAK, Gérard PANCZER, Florian TRICHARD, Bruno BOUSQUET, Frédéric PELASCINI a Vincent MOTTO-ROS: *Exploration of megapixel hyperspectral LIBS images using principal component analysis*. Journal of Analytical Atomic Spectrometry [online]. 2018, 33(2), 210-220 [cit. 2019-05-06]. DOI: 10.1039/C7JA00398F. ISSN 0267-9477. Dostupné z: <http://xlink.rsc.org/?DOI=C7JA00398F>
- [37] SHRESTHA, Leбина, Konstantina VIOLAKI: *Line Broadening [online]*. Department of Medical Physics and Bioengineering University College London. 2011 [cit. 2019-05-02]. Dostupné z: <https://www.youtube.com/watch?reload=9&v=Dvk0J0jx-Uk&fbclid=IwAR3jcKf81Z6FVqU-dImkjpT-wDIIdYbiYN5qzfWuo8GcR9cZBcj5EUM1FVwk>.
- [38] KRAMIDA, A., Yu. Ralchenko , J. Reader and NIST ASD Team (2017): *NIST Atomic Spectra Database (version 5.5.1)*, [Online]. . Dostupné z:<https://physics.nist.gov/asd> [11.10.2018] National Institute of Standards and Technology, Gaithersburg, MD.

Přílohy

Přílohy jsou na CD přiloženém k práci. Tyto přílohy obsahují zdrojová data a programy vytvořeny v software Python. Je zde zdrojový kód a také vyexportovaný soubor ve formátu .html, vhodný pouze pro čtení. Pro .html soubory je nutné mít připojení k internetu.

Příloha obsahuje:

- A) PCA 6 prvků
- B) PCA 6 prvků redukce50
- C) PCA 2 prvky přechod
- D) PCA 2 prvky otutliery
- E) PCA 2 prvky 3Fe+3Al
- F) PCA 2 prvky Fe+3Al aproximace
- G) Předzpracování
- H) Zdrojová data