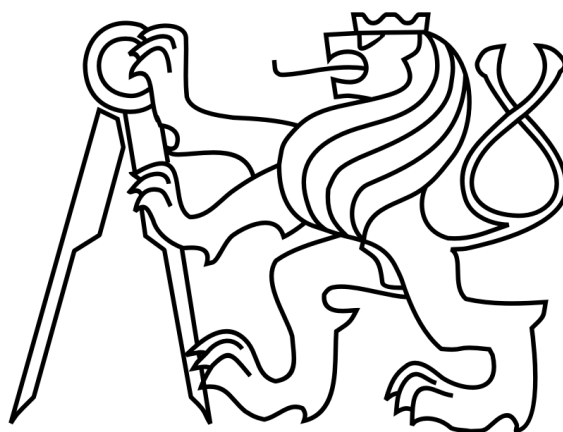


ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE

FAKULTA ELEKTROTECHNICKÁ

KATEDRA RADIOELEKTRONIKY



DIPLOMOVÁ PRÁCE

Detekce mraků podle barevné informace z celoblohových kamer

Bc. Lukáš Krauz

Vedoucí práce: Mgr. Martin Blažek, Ph. D.

Praha, 2019

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Krauz** Jméno: **Lukáš** Osobní číslo: **420246**
Fakulta/ústav: **Fakulta elektrotechnická**
Zadávající katedra/ústav: **Katedra radioelektroniky**
Studijní program: **Elektronika a komunikace**
Studijní obor: **Audiovizuální technika a zpracování signálů**

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

Detekce mraků podle barevné informace z celooblohových kamer

Název diplomové práce anglicky:

Cloud Detection Based on Colour Space Information from All-sky Cameras

Pokyny pro vypracování:

Zpracujte data z celooblohové kamery WILLIAM. Z těchto dat vytvořte databázi typů mraků podle jejich pozice v barevných prostorech La^*b^* a xyz. K rozlišení různých oblastí při zpracování obrazu využijte metodu K-means segmentace. Analyzujte data v závislosti na proměnlivých denních a světelných podmínkách. Pomocí analýzy hlavních komponent zkuste najít vzájemné závislosti v získaných datech. Proveďte hloubkovou analýzu těchto závislostí a navrhněte optimální kritéria pro statistické rozlišení různých typů mraků v barevných prostorech včetně analýzy chyb a pravděpodobností. Diskutujte vhodnost této metody pro využití v kombinaci s dalšími meteorologickými přístroji.

Seznam doporučené literatury:

- [1] Blažek M., Pata P.: Colour transformations and K-means segmentation for automatic cloud detection, Meteorologische Zeitschrift, Vol. 24, No. 5, 503–509 (2015).
- [2] P. Janout, P. Páta, Celooblohová kamera s extrémně širokouhlým zorným polem, Jemná mechanika a optika, 2014, 59(10).
- [3] Gonzales R. C., Woods R. E., Digital Image Processing, 3. vydání, Pearson, 2008.

Jméno a pracoviště vedoucí(ho) diplomové práce:

Mgr. Martin Blažek, Ph.D., Instituto de Astrofísica de Andalucía, CSIC

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) diplomové práce:

Datum zadání diplomové práce: **13.09.2018**

Termín odevzdání diplomové práce: **08.01.2019**

Platnost zadání diplomové práce: **19.02.2020**

Mgr. Martin Blažek, Ph.D.
podpis vedoucí(ho) práce

podpis vedoucí(ho) ústavu/katedry

prof. Ing. Pavel Ripka, CSc.
podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

Diplomant bere na vědomí, že je povinen vypracovat diplomovou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v diplomové práci.

Datum převzetí zadání

Podpis studenta

Abstrakt

Tato diplomová práce se věnuje návrhu detekčního algoritmu atmosferických jevů (mraků) podle barevné informace z celooblohových kamer. Snímání denní oblohy pomocí širokouhlých celooblohových systémů je v meteorologii čím dál více aplikováno a možnost využití barevné obrazové informace za účelem rozboru denní oblohy byla taktéž v minulosti potvrzena. Tato diplomová práce na uvedených předpokladech staví. K vyvinutí funkčního algoritmu detekce mraků je zapotřebí signifikantního množství obrazových dat. Popis jejich získávání, anotace, uložení do databáze, segmentace a následná analýza v různých barevných prostorech pracující s těmito daty je jádrem této práce. Pro detailní rozbor barevné informace mraků po dobu celého dne jsou zde představeny solární parametry, na jejichž změně byl chromatický vývoj zkoumán. Kvůli multidimenzionalitě získaných závislostí je v této práci taktéž testována statistická metoda hlavních komponent. Vyhodnocení analýzy zpracovaných dat dovolilo navrhnout kritéria pro algoritmus detekce a klasifikace oblaků, která jsou zde založena na představených závislostech barevné informace a její vhodné reprezentace ve zvolených barevných prostorech. Proces celého algoritmu je v této diplomové práci rozdělen na metody klasifikace pracující se zmíněnými kritérii, metody předzpracování obrazu a funkce zajišťující výstupní data. S těmito výstupními daty je zároveň představena možnost detekce srážkové oblačnosti spolu s vytvořeným indikátorem varujícím před tímto jevem. Aplikace a funkčnost algoritmů vyvinutých v této diplomové práci byla testována a ověřena na obrazových datech v reálném provozu celooblohového systému WILLIAM.

Abstract

This thesis summarizes development of atmospheric phenomena (clouds) detection algorithm, based upon colour information from all-sky cameras. Daytime sky evaluation via a widefield whole-sky imaging systems is nowadays in meteorology very common and also the utilization of chromatic information to detect clouds has been introduced. All research in this thesis uses these premises. For creation of successful cloud detection algorithm is vital to acquire a significant amount of imaging data. The description of its acquisition, annotation, saving to database, segmentation and further analysis in different colour spaces takes a great part in this thesis. For a detail research of clouds colour information during the day, solar parameters are introduced and based upon its values other chromatic features analysed. Because of the multidimensionality of acquired data the principal component analysis is introduced as well. The evaluation of overall analysis in this thesis allowed to propose various criteria for the main cloud detection and classification algorithm. Proposed algorithm is divided in different methods dealing with classification, image pre-processing and managing the output data. This data allowed to propose rules for indication of raining clouds as well. The developed method for cloud detection and classification has been tested and verified on a real-time data from wide-field all-sky image analyzing system WILLIAM.

Klíčová slova: Barevné prostory, klasifikace oblaků, meteorologie, RGB, CIELAB, CIEXYZ, lineární regrese, křížová validace, PCA, WILLIAM, K-means, DSLR, celoblohový systém, detekce oblaků.

Keywords: Color spaces, cloud classification, meteorology, RGB, CIELAB, CIEXYZ, linear regression, PCA, cross-validation, WILLIAM, K-means, DSLR, whole sky imaging system, cloud detection.

Poděkování

Tímto bych rád poděkoval mému vedoucímu Mgr. Martinu Blažkovi, Ph.D. za hodnotné připomínky k mé diplomové práci a jeho drahocenný čas při konzultacích. V neposlední řadě děkuji i své rodině, partnerce, přátelům a lidem mně blízkým za trpělivost, pomoc při korekci a celkovou podporu.

Prohlášení

Prohlašuji, že jsem svou diplomovou práci „Detekce mraků podle barevné informace z ce-
looblohových kamer“ vypracoval samostatně a uvedl veškeré použité informační zdroje
v souladu s Metodickým pokynem č. 1/2009 o dodržování etických principů při přípravě
vysokoškolských závěrečných prací.

V Praze dne 7. ledna 2019,

.....

Podpis studenta

Obsah

1	Úvod	17
2	Celooblohová kamera WILLIAM	18
2.1	Pozorovací stanice.....	18
2.2	Konstrukce	19
2.3	Software.....	20
2.4	Datový výstup	20
3	Meteorologická klasifikace a barevné prostory	23
3.1	Rozdělení oblaků	23
3.1.1	Třídy rozdělení oblaků	23
3.2	Vybrané typy barevných prostorů	23
3.2.1	CIE 1931 XYZ	24
3.2.2	CIELAB	24
3.3	Základní statistické pojmy	25
3.3.1	Střední hodnota a rozptyl	25
3.3.2	Kovariance.....	26
3.3.3	Korelace	27
3.4	Barevná segmentace pomocí K-means shlukování	27
4	Zpracování dat z celooblohového kamerového systému WILLIAM	29
4.1	Předzpracování snímků	29
4.2	Vytvoření databáze a anotací kamerových snímků	30
4.3	Solární parametry	31
4.3.1	Pravé poledne	32
4.3.2	Elevace slunce	32
4.3.3	Časová vzdálenost od východu / západu slunce	33
4.4	Selekce použitelných dat	33

5	Analýza meteorologických jevů	36
5.1	Barevné diagramy mraků	36
5.2	Analytický popis barevných diagramů	41
5.2.1	Regresní analýza	42
5.2.2	Cross-validation	44
5.2.3	Predikční intervaly	46
5.2.4	Výsledky regresní analýzy	46
5.3	Analýza hlavních komponent	52
5.3.1	Představení analýzy hlavních komponent	52
5.3.2	Výsledky PCA pro solární parametry	54
6	Užití algoritmu pro včasnou výstrahu při předpovědi počasí.....	55
6.1	Algoritmus detekce a klasifikace oblaků	55
6.2	Kritéria hodnocení pravděpodobnosti	57
6.2.1	Pravděpodobnost dle závislosti složky b^* a času od vý- chodu / západu slunce	57
6.2.2	Pravděpodobnost dle závislosti složky b^* na elevaci slunce	58
6.2.3	Pravděpodobnost dle vzdálenosti od chybových elips v prostoru a^* , b^* , respektive L , b^*	58
6.3	Automatická klasifikace v praxi	60
7	Závěr	65
	Seznam použitých zkratk	67
	Seznam obrázků	68
	Seznam tabulek	69
	Literatura	70
	Příloha A	76
	Příloha B	79

1 Úvod

Tato diplomová práce navazuje na předchozí výzkum a analýzu denní oblohy v původní bakalářské práci [1] zabývající se barevnou reprezentací atmosferických jevů, zejména oblaků. Zároveň tak vylepšuje a podstatně rozšiřuje možnosti projektu celooblohové kamery WILLIAM [2], který je provozován katedrou Radioelektroniky na fakultě Elektrotechnické a který se zaměřuje na evaluaci povětrnostních podmínek. Veškerým podrobnostem uvedeného projektu se věnuje následující kapitola.

Tvorba celooblohových snímacích systémů pro hodnocení stavu oblohy dnes není žádnou novinkou a různé výzkumné skupiny se napříč celým světem tímto problémem zabývají. Kromě astronomického snímání hvězd, které zde kvůli odlišnému námětu této práce opomineme, tyto systémy velice často detekují oblačnost a jsou pro ně vyvíjeny nejrůznější algoritmy založené na různorodých předpokladech pro tento jev. Zde můžeme uvést například projekty detekce a klasifikace oblaků ve vysokém dynamickém rozsahu, snímáním oblohy v IR (Infra Red) oblasti a jiné metody detekce oblaků [3–8]. Část z nich se zabývá obrazovou segmentací za pomoci barevného rozlišení získaných snímků. Dalšími jsou například [9, 10] nebo články zaměřené na spektrální analýzu denního světla [11, 12].

V této práci vyvíjíme metodu barevných diagramů za pomoci k-means segmentace představené v [13]. V mé předchozí bakalářské práci [1] bylo dokázáno, že po určitém rozdělení mraků do podobnostních skupin je možné v různorodých barevných prostorech tyto mraky jak detekovat, tak i klasifikovat.

V dalších kapitolách budou některé prvky mé předchozí bakalářské práce [1] postupně rozšířeny pro mnohem větší objem získaných dat. Nově v této diplomové práci bude uvedena obsáhlejší analýza nalezených skupin oblaků v různých barevných reprezentacích, jejich chování a závislosti na nově zavedených pozorovacích parametrech. Bude zde také možné ukázat rozdílnost ve vlastnostech jednotlivých oblaků, respektive jejich tříd v průběhu denního cyklu.

Závěrečná kapitola 6 představuje zcela nový algoritmus pro detekci, klasifikaci a určení dešťové oblačnosti. Ten je konstruován pro včasné upozornění na srážky, a to po vyhodnocení objektů (prakticky vždy oblaků) nacházejících se na snímcích denní oblohy. Výstupy tohoto algoritmu budou v poslední kapitole předvedeny na ukázce dat nebeské oblačnosti z pozorovací stanice v Jarošově nad Nežárkou.

Popis nových výpočetních skriptů, ať již z prostředí MATLAB, které bylo pro analýzu oblaků využíváno, tak v programovacím jazyce Python, ve kterém byly veškeré funkce pro představený automatický algoritmus naprogramovány, je uveden v přílohách na konci této diplomové práce. K němu jsou přidány i nové ukázky vstupů a výstupů, které byly v této práci vytvořeny. Kompletní naprogramovaný kód je k nalezení na přílohovém CD.

2 Celoblohová kamera WILLIAM

Vývoj algoritmů pro klasifikaci oblačnosti byl od základu tvořen pro širokoúhlý systém WILLIAM, celým názvem (Wide-field aLL-sky Image Analyzing Monitoring system) [2]. Jedná se o projekt navržený Katedrou radioelektroniky, Fakulty elektrotechnické, ČVUT v Praze, zabývající se celoblohovým snímáním kamerou s extrémně širokoúhlým zorným polem, která umožňuje monitorování atmosferických jevů denní i noční oblohy, detekci hvězd a celkovou analýzu povětrnostních podmínek [14–16].

Tato diplomová práce je založena primárně na analýze snímků denní oblohy, čemuž byla přizpůsobena i veškerá výstupní data. Možnost využití barevných prostorů pro klasifikaci oblaků za denního osvětlení byla popsána v [13]. V návaznosti byl uvedený článek rozšířen pro celoblohové kamerové systémy v bakalářské práci [1]. Primární cíl, pro který byl systém WILLIAM vyvíjen, však tvořilo pořizování ultra-širokoúhlých záznamů noční oblohy a na těchto snímcích vyhodnocovat pohyb nebeských objektů, mimo jiné hvězd a komet [16–18]. Výčet dnes zachytitelných atmosferických a astronomických jevů je již poměrně dlouhý, a díky tomu je možné tento systém upravovat pro širokou škálu dalších aplikací.

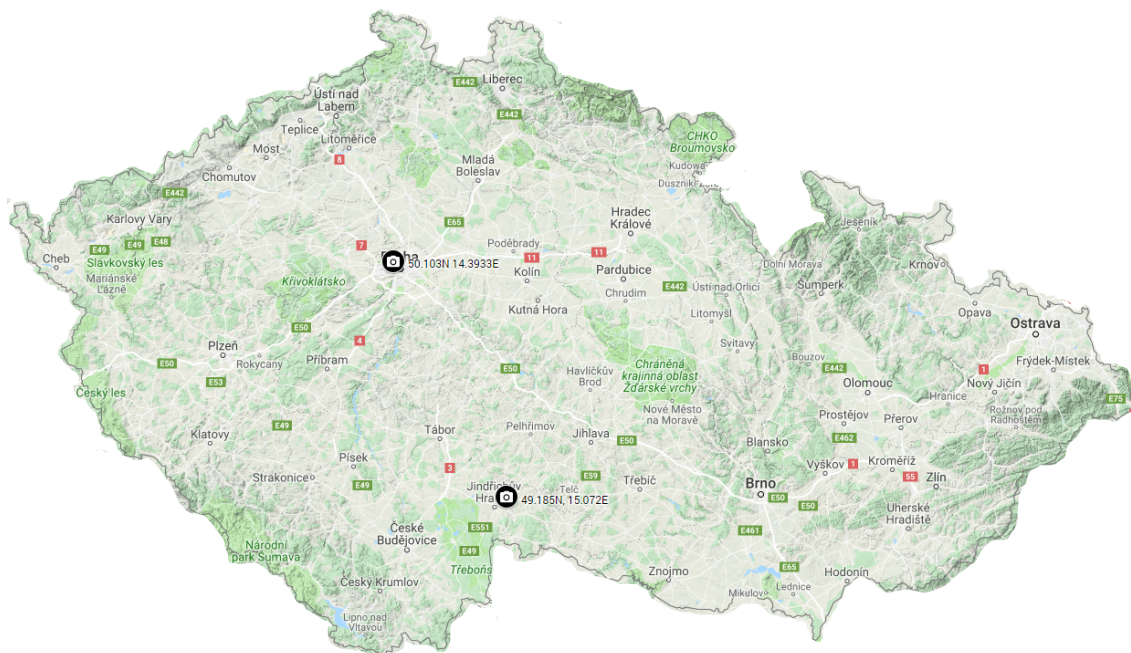
Neustále se zvětšující databáze extrémně-širokoúhlých snímků nalézají také různá využití při tvorbě nových algoritmů pro zpracování obrazu a umožňuje tyto algoritmy na zmiňovaných záznamech testovat. WILLIAM byl využit pro tvorbu postupů výpočtu PSF (Point Spread Function) širokoúhlých systémů [19], a následně jejich modelováním a popisem [20,21]. Mezi další publikace, které se zabývaly nebo nějakým způsobem využívaly celoblohovou kameru WILLIAM můžeme řadit i [22,23] věnující se astrometrické kalibraci extrémně-širokoúhlých kamerových systémů nebo například dekonvolučnímu potlačení rozostření snímků získaných z těchto systémů.

2.1 Pozorovací stanice

Systém WILLIAM byl od prvopočátku zamýšlen ve více verzích nacházejících se na různých bodech napříč Českou republikou s možností detekovat co největší plochu oblohy rozprostírající se nad tímto územím. Díky tomu by bylo možné identifikovat mnoho atmosferických a vesmírných jevů na nebi po dobu celého dne. V budoucnu, v případě nastavení snímání v oblasti IR, existuje možnost zaznamenávat noční pohyb hvězd i za špatné viditelnosti při vysoké oblačnosti. Nevýhodou však může být komplikovaný návrh optiky a snímacích zařízení.

Momentálně WILLIAM funguje ve 2 verzích (obr. 1). Hlavní instalace, postavená na typu kamery DSLR (Digital Single Lens Reflex Camera), se od června 2015 nachází v jižních Čechách v obci Jarošov nad Nežárkou (GPS souřadnice 49.185N, 15.072E). Hlavním z důvodů pro volbu této lokace bylo především nízké světelné znečištění, vzdálenost od dalších pozorovacích stanic a vhodná přístupnost k systému.

Druhá verze celoblohového systému WILLIAM se nachází v Praze na střeše budovy Fakulty elektrotechnické, ČVUT v Praze (GPS souřadnice 50.103N 14.3933E). Zde instalace funguje již od poloviny roku 2014 a od roku 2017 s novým přebudovaným systémem, založeným na snímání astronomickou kamerou společnosti ZWO, představené dále.



Obrázek 1: Umístění pozorovacích stanic systému WILLIAM v České republice. První se nachází v Praze a druhá v Jarošově nad Nežárkou u Jindřichova Hradce.

2.2 Konstrukce

Konstrukce systému zaznamenávajícího obrazovou informaci je povětšinou založena na snímajících přístrojích respektive kamerách nebo fotoaparátech. WILLIAM dnes využívá 2 druhy takovýchto zařízení. Prvním byl DSLR fotoaparát NIKON D5100¹ [24] se snímačem s architekturou CMOS (Complementary Metal Oxide Semiconductor) o velikosti 23,6x15,6 mm, vytvářející rozlišení 4928x3264 pixelů, který však byl roku 2017 nahrazen typem NIKON D5300² se stejně velkým CMOS čipem, ale s velikostí obrazu 6000x4000 pixelů. Druhým zařízením je již zmiňovaná astronomická kamera ZWO ASI 178MC³ s rozlišením 3096x2080 obrazových bodů a rovněž technologií CMOS.

Vývoj nových generací systému WILLIAM počítá i s osazením větších kamer vhodných pro astronomické snímání. V případě plánované kamery ASI 1600MM-Cool⁴ však nastává problém s barevnou detekcí mraků (na které je založena tato práce). Jelikož se jedná o kameru monochromatickou, bude muset být proces interpolace pomocí Bayerovské masky [25] nahrazen např. přídatným barevným filtrem nebo jiným postupem [14].

K celooblohovému snímání je důležité nasadit na fotoaparát odpovídající objektiv zabírající co největší zorné pole. Aktuálně se jedná o dva různé druhy. Pro WILLIAM v Jarošově je použit objektiv typu rybí oko [1, 26] Sigma 10mm F2.8 EX DC 2⁵, dále popsán zde [1], a pro pražskou instalaci širokoúhlý objektiv ZWO CCTV Lens 2.5mm 170 Degree⁶.

¹<https://imaging.nikon.com/lineup/dslr/d5100/>

²<https://imaging.nikon.com/lineup/dslr/d5300/>

³<https://www.astroshop.eu/astronomical-cameras/zwo-camera-asi-178-mc-color/p,48568>

⁴<https://astronomy-imaging-camera.com/product/asi1600mm>

⁵<https://www.sigmaphoto.com/10mm-f2-8-ex-dc-hsm-fisheye>

⁶<https://astronomy-imaging-camera.com/product/zwo-2-5mm-150-degree-lens>

Zmiňované komponenty jsou spolu s napájením osazeny ve vyhřívaném držáku kloubového charakteru s průhlednou vrchní kopulí, který podle návrhu musí být voděodolný, neprodyšný a uzpůsobený pro osazování různými typy kamer. Kontrolu nad tímto stavem udržuje variace senzorů, např. teploty, tlaku či vlhkosti. Takto navržený systém dokáže odolat nepříznivým až kritickým povětrnostním podmínkám bez známek poruchy funkčnosti.

Řízení systému a GPS kontrolu přesného denního času provádí autonomně Raspberry Pi 2 a 3⁷, které je součástí každé z pozorovacích stanic. Veškeré následné výpočty však probíhají na vzdáleném serveru. Navržený proces je zjednodušeně popsán v další podkapitole.

2.3 Software

Jak bylo uvedeno výše, Raspberry Pi u každé verze systému WILLIAM umožňuje autonomní řízení. Každé z nich tak využívá primárně nastavený operační systém Raspbian⁸ (odvozen z nejstarší GNU/Linuxové distribuce Debian⁹). Zde jsou připravené naprogramované skripty spouštěné softwarovou utilitou Cron¹⁰ typickou pro linuxové operační systémy. Raspberry následně komunikuje se vzdáleným serverem, na který jsou odeslána veškerá data a poté prováděny výpočty. Přístup na tento server je zprostředkován například pomocí šifrovaného SSH (Secure Shell) protokolu, případně k datům můžeme použít i přístup pomocí FTP (File Transfer Protocol) protokolu.

Jednotlivé skripty určující chod a aplikaci algoritmů systému mohou být vytvářeny v různých programovacích jazycích. Důležitým bodem je uvedení interpretovaného jazyka Python¹¹, respektive jeho verzi Python 3.0¹² a vyšší, který byl využíván pro aplikaci navrženého algoritmu klasifikace oblaků (představen v dalších kapitolách), a grafický, příkazovou řádkou řízený program Gnuplot¹³, využitelný na téměř všech dnes známých platformách.

Schéma softwarové části systému WILLIAM je zobrazeno na obrázku 2, a značí jednotlivé funkční bloky určující chod celého systému. WILLIAM také generuje výstupy, podle toho v jakém režimu pracuje, a je velice důležité rozlišit, zda se jedná o zpracování informace týkající se oblačnosti nebo zda snímá oblohu za účelem detekce hvězd. S oběma režimy však navržený software dokáže pracovat.

2.4 Datový výstup

Jako autonomní snímací kamerový systém umí WILLIAM exportovat získané fotografie v různých datových formátech. Většinou jsou vzniklé snímky ukládány v RAW formátu s JPEG (Joint Photographic Experts Group) náhledy, případně ve formátech NEF (Nikon Electronic Format) typických pro DSLR od značky NIKON. Ty jsou vždy vytvořené jako surová data přicházející přímo ze snímače fotoaparátu a lze je následně jakkoliv upravovat a transformovat.

⁷<https://www.raspberrypi.org/>

⁸<https://www.raspberrypi.org/downloads/raspbian/>

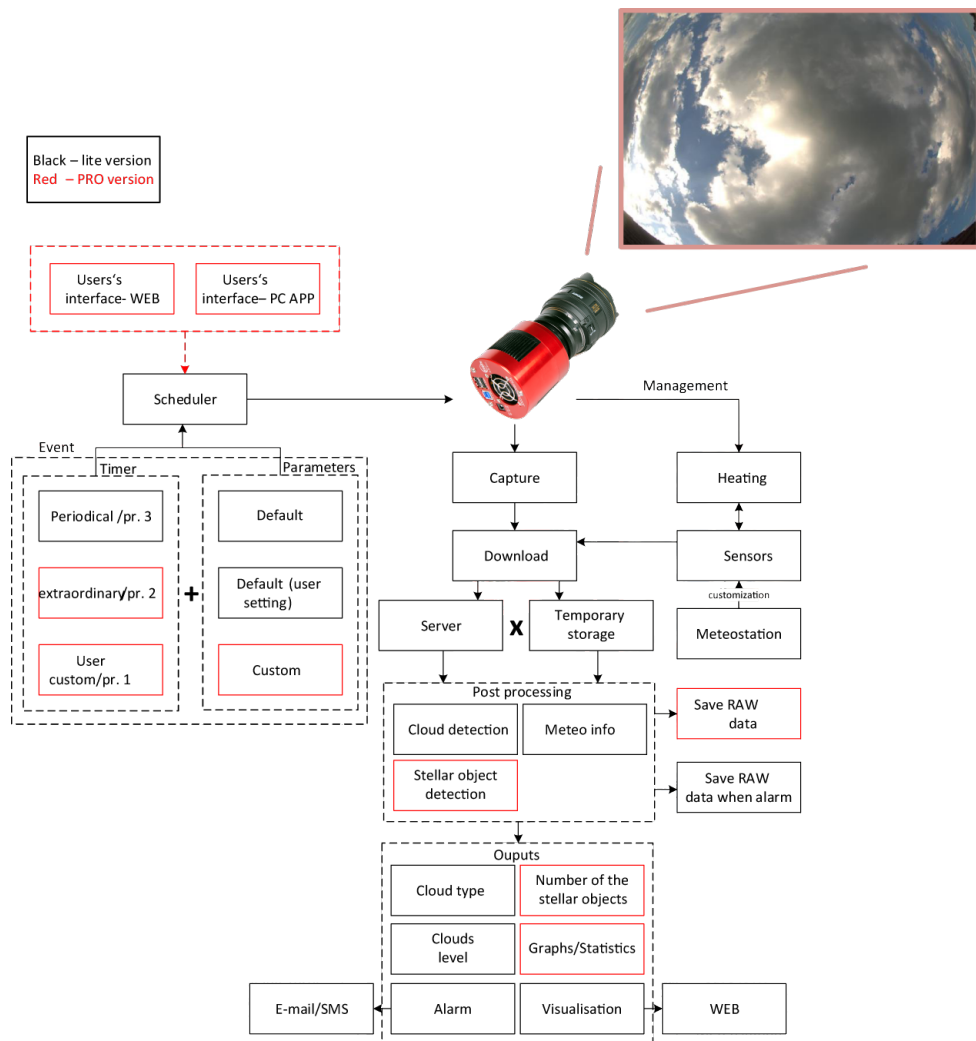
⁹<https://www.debian.org/>

¹⁰<https://help.ubuntu.com/community/CronHowto>

¹¹<https://www.python.org/>

¹²<https://www.python.org/download/releases/3.0/>

¹³<http://www.gnuplot.info/>

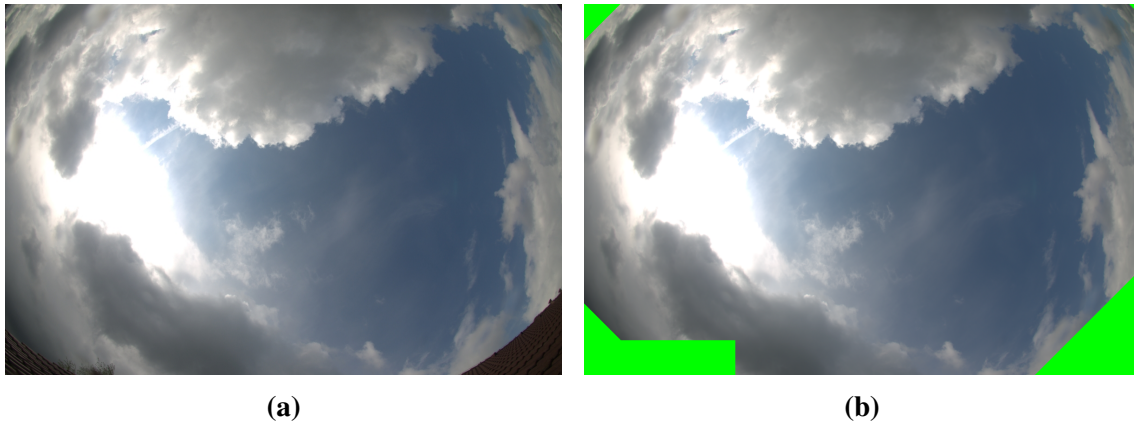


Obrázek 2: Schéma funkcionality systému WILLIAM společně s připravovaným rozšířením. Na obrázku jsou znázorněny procesy spravování obrazové informace, veškerých dat a výstupu, které systém produkuje¹⁴ [14].

Jelikož se při snímání denní oblohy mohou světelné podmínky po dobu celého dne měnit, je pro detekci mraků nutné sjednotit parametry fotografií do shodného podání barev. Z těchto důvodů bylo přistoupeno k zafixování a vyvážení barevné teploty bílé barvy na 6000K a následnému transformování do bezztrátového obrazového formátu TIFF (Tagged Image File Format). Tato konverze může být například prováděna externě z příkazové řádky ve volně dostupném softwaru, zaměřujícím se na práci s RAW daty, nebo často přímo automaticky ve fotoaparátu.

Pro barevnou analýzu mraků na obloze je nutné odstranit, nebo lépe zamaskovat, nechtěné statické objekty objevující se v expozici (obr. 3a). Statické objekty se mohou při barevné segmentaci nechtěně spojovat s některými druhy oblaků nebo tvořit svůj vlastní cluster. Proto byla v předchozí bakalářské práci [1] navržena barevná pixelová maska, zakrývající tyto statické elementy (obr. 3b). Ta je aplikována pro každý snímek vstupující ke zpracování do algoritmu.

S principem předzpracování vstupních snímků pro popisovaný algoritmus také souvisí, že není nutné používat fotografie s RAW rozlišením 16 a více megapixelů, které produkují



Obrázek 3: Na obrázku (a) je ukázán snímek denní oblohy, v měsíci květnu, získaný ze systému WILLIAM v Jarošově nad Nežárkou. Druhý obrázek demonstruje použití barevné masky pro odstranění statických objektů ze snímku, které narušují další „post-processing“ prováděný na těchto fotografiích.

fotoaparáty zabudované v systému WILLIAM a obsahují velké množství redundantní informace. Kvůli tomu jsou data podvzorkována na zvolených 800x1200 pixelů [1], tím se zároveň sníží výpočetní náročnost dalšího zpracování a algoritmus tak lze nasadit na vyhodnocování snímků v reálném čase.

3 Meteorologická klasifikace a barevné prostory

V článku [13] o meteorologické klasifikaci byla představena možnost segmentovat jednotlivé typy mraků podle svých barevných podobností. Tato skutečnost a navržený algoritmus barevné segmentace byl později ověřen a prokázán i pro celoblohové kamery v mé bakalářské práci [1]. Výsledkem tohoto byl fakt, že jednotlivé mraky, rozdělené do skupin podle strukturálních a vizuálních podobností, se spojují do určitých clusterů ve zvolených barevných prostorech za pomoci segmentačního algoritmu k-means. Vzhledem k tomu, že tato diplomová práce na uvedený experiment navazuje, je nutné znovu uvést některé základní pojmy.

3.1 Rozdělení oblaků

Meteorologický rozbor oblaků je podrobně popsán v [1], kde byly představeny jednotlivé druhy oblačnosti [27, 28], jejich rozdělení podle výšky základny, vzhledu a tvaru. V této diplomové práci použijeme stejné rozdělení do elementárních skupin (tříd) na základě vizuální i fyzikální podobnosti těchto mraků, které bylo nutné kvůli funkčnosti algoritmu pro klasifikaci mraků vytvořit. Pojmenování a bližší informace o existujících druzích oblačnosti je možno nalézt v [1, 27, 28].

3.1.1 Třídy rozdělení oblaků

Toto rozdělení do tříd oblaků bylo vytvořeno již v [1] a v této práci z něj dále vycházíme, neboť se zároveň jedná o hlavní třídy využitě pro klasifikaci nových dat.

1. Cirrus, Cirrostratus, Cirrocumulus, Altopcumulus
2. Altopcumulus, Cumulus, Stratocumulus, (Cumulonimbus)
3. Altostratus, Stratus, Cumulonimbus, Nimbostratus
4. čistá obloha

Jedním z důležitých parametrů pro určení tříd oblaků bylo indikování srážkovosti. Oblaka srážková jsou především ve třídě 3, potažmo částečně ve třídě 2. Jedná se o **Stratocumuly** a **Cumuly**, ovšem ne všech typů [27], v jistých případech **Stratus** a **Altostratus** [28], a především bouřkové **Cumulonimbus** a **Nimbostratus**. Ve zbylých případech se nachází oblaka bezesrážková.

3.2 Vybrané typy barevných prostorů

Barevnou informaci na snímcích denní oblohy je za určitých podmínek možné rozsegmentovat [1, 13]. K tomu ovšem potřebujeme správnou reprezentaci barevných hodnot, především v barevných prostorech k této úloze určených. Zvláště nevhodné jsou základní barevné prostory RGB, respektive sRGB, obsahující ve všech svých složkách informaci o jasů. Z tohoto důvodu je nutné přistoupit k prostorům, které tuto informaci separují. Dle [13] se jeví jako nejvhodnější prostory CIELAB [29] a klasický CIE 1931 XYZ [30]. Podrobnosti o dalších barevných prostorech jsou k nalezení v [1], zde pouze pro úplnost uvedeme zmíněné 2 nejdůležitější.

3.2.1 CIE 1931 XYZ

Pro potřebu vyjádření všech barev reálnými nezápornými čísly bylo z modelu sRGB, často používané podmnožiny RGB prostoru reprezentující skutečné barvy, ustoupeno. Nově však vzniká prostor CIE XYZ s barvami neskutečnými a to již v roce 1931. Ten vychází z měření citlivosti lidského oka a ze zavedení pojmu standardní pozorovatel [31]. Jednotlivé složky prostoru XYZ vychází z lineární transformace r, g, b kanálů [32]. Přepočítání ze základního barevného prostoru lze vyjádřit následujícím maticovým zápisem

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} x_R & x_G & x_B \\ y_R & y_G & y_B \\ z_R & z_G & z_B \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \quad (3.2.1.1)$$

Každý z koeficientů $x_{R,G,B}$, $y_{R,G,B}$, $z_{R,G,B}$ má hodnoty podle standardního zdroje denního osvětlení, např. D65, D50 atd [33]. Zde používaný převod do CIE XYZ prostoru případně i do CIELAB (viz kapitola 3.2.2) využívá základní barevnou teplotu zdroje D65 odpovídající dennímu osvětlení oblohy. Pro tento případ můžeme uvést hodnoty zmíněných koeficientů v tabulce 1.

Tabulka 1: Matice koeficientů pro převod z sRGB do XYZ pro zdroj světla D65

$$\begin{bmatrix} 0.4124564 & 0.3575761 & 0.1804375 \\ 0.2126729 & 0.7151522 & 0.0721750 \\ 0.0193339 & 0.1191920 & 0.9503041 \end{bmatrix}.$$

Vzniklé komponenty X, Z byly v tomto případě voleny, aby představovaly určitou základní barvu s daným tónem a sytostí, avšak bez informace o jasu. Ten je zcela reprezentován třetí komponentou Y. CIE XYZ je zároveň navržen tak, aby jednotkové hodnoty těchto komponent tvořily bílé světlo. Přepočítání na pouze barevné hodnoty bez jasů pro zobrazení do tzv. MKO normalizované soustavy je následující

$$x = \frac{X}{X + Y + Z}, \quad y = \frac{Y}{X + Y + Z}, \quad z = \frac{Z}{X + Y + Z}. \quad (3.2.1.2)$$

Pro kolmou souřadnicovou soustavu stačí pouze souřadnice x , y , zbylá z lze dopočítat pomocí vztahu

$$x + y + z = 1, \quad (3.2.1.3)$$

jelikož se jedná o jejich lineární kombinaci. Takto je vytvořen tzv. barevný prostor Yxy, kdy složky x , y odpovídají předchozím složkám ze vztahu 3.2.1.2, Y je komponenta vycházející z rovnice 3.2.1.1 a, jak bylo uvedeno, obsahuje informaci o jasu.

3.2.2 CIELAB

Hlavní místo mezi barevnými prostory využívanými pro algoritmus klasifikace mraků zaujímá nezávislý prostor CIELAB, pro který bylo ukázáno v [1], že jednotlivá oblaka lze barevně nejlépe odlišit při různém zavedení závislostí komponent tohoto prostoru. Velkou devizou je možnost oddělitelnosti jasové složky a zanechání pouze barevné informace

ve dvou zbylých chromatických komponentách, jako tomu bylo v prostoru CIE XYZ, který tvoří jeho základ, avšak není tolik perceptuálně uniformní. Každou komponentu jde poměrně lehce transformovat z původních trichromatických složek CIE XYZ prostoru následujícími rovnicemi (převzato z [1])

$$L = 116 \left(\frac{Y}{Y_n}^{\frac{1}{3}} - 16 \right), \quad (3.2.2.1)$$

$$a = 500 \left[\left(\frac{X}{X_n} \right)^{\frac{1}{3}} - \left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} \right], \quad (3.2.2.2)$$

$$b = 200 \left[\left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} - \left(\frac{Z}{Z_n} \right)^{\frac{1}{3}} \right], \quad (3.2.2.3)$$

pro $X > 0, 008856X_n, Y > 0, 008856Y_n, Z > 0, 008856Z_n$, respektive

$$L = 903,33 \left(\frac{Y}{Y_n} \right)^{\frac{1}{3}} - 16, \quad (3.2.2.4)$$

$$a = 7,787 \left[\frac{X}{X_n} + \frac{16}{116} \right] - 7,787 \left[\frac{Y}{Y_n} + \frac{16}{116} \right], \quad (3.2.2.5)$$

$$b = 7,787 \left[\frac{Y}{Y_n} + \frac{16}{116} \right] - 7,787 \left[\frac{Z}{Z_n} + \frac{16}{116} \right]. \quad (3.2.2.6)$$

při nerovnosti $X \leq 0, 008856X_n, Y \leq 0, 008856Y_n, Z \leq 0, 008856Z_n$. Konstanty X_n, Y_n, Z_n znázorňují trichromatické složky normalizovaného světla a například pro světelný zdroj D65 mají hodnoty uváděné v tabulce 2.

Tabulka 2: X_n, Y_n, Z_n koeficienty CIELAB prostoru pro zdroj světla D65

X_n	Y_n	Z_n
95.047	100	108.883

3.3 Základní statistické pojmy

Pro analýzu a další zpracování vytvořených dat, které budou popsány v dalších kapitolách, je pro úplnost nutné v krátkosti uvést základní statistické pojmy a výpočty, které byly v této práci aplikovány a na které se budeme dále odkazovat.

3.3.1 Střední hodnota a rozptyl

Střední hodnota, v tomto případě zamýšlená jako aritmetický průměr [34], je uváděna při měření diskrétních hodnot jedné proměnné X jako

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x_i, \quad (3.3.1.1)$$

kde μ_x nebo také \bar{x} značí aritmetický průměr, x_i přísluší jednotlivým vzorkům proměnné X , N reprezentuje počet vzorků. Střední hodnota tak říká, kde jsou jednotlivá měření proměnné X soustředěny.

Pro statistické hodnocení je velice důležitý i další parametr, a to zjištění rozptylu dat proměnné X , neboli variance, vzorcem definovaným jako

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2. \quad (3.3.1.2)$$

Rozptyl tedy udává výchylku dat od střední hodnoty. Odmocnina tohoto parametru je známá také pod názvem standardní odchylka, vztahem reprezentována jako

$$\sigma = \sqrt{\sigma^2}, \quad (3.3.1.3)$$

neboli

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2}, \quad (3.3.1.4)$$

kde σ značí standardní odchylku, μ_x odpovídá aritmetickému průměru, x_i jsou jednotlivé vzorky proměnné X a N počet vzorků. Jelikož ve většině případů nemáme úplný základní soubor hodnot, musí být zde uváděné statistické parametry odhadovány s počtem stupňů volnosti, který je roven $N - 1$.

3.3.2 Kovariance

Zkoumáme-li 2 různé veličiny, např. X , Y , můžeme hledat statistické závislosti mezi nimi, čímž se budeme zabývat především v kapitole 5. Výpočet této závislosti lze vyjádřit jako

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y). \quad (3.3.2.1)$$

σ_{xy} zde zastává kovarianci, μ_x, μ_y jsou aritmetické průměry veličin X, Y , jejich vzorky jsou x_i, y_i a N opět počet těchto vzorků [35].

Zároveň platí vztah 3.3.2.2 značící, že rozptyl může být kovariancí sama sebe pro jednu proměnnou veličinu X .

$$\sigma^2 = \sigma_{xx}. \quad (3.3.2.2)$$

Výsledkem při výzkumu statistických závislostí 2 veličin X, Y dostáváme symetrickou kovarianční matici \mathbf{K} reprezentující vztah 3.3.2.3 s elementy rozptylu jednotlivých veličin na diagonále.

$$\mathbf{K} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix}. \quad (3.3.2.3)$$

Hodnoty, respektive znaménka jednotlivých vzájemných kovariancí udávají, jakým způsobem a směrem jsou na sobě jednotlivé veličiny závislé. Míru jejich závislosti však mnohem lépe popisuje pojem korelace.

3.3.3 Korelace

Korelace [35] jako parametr pracuje s významem kovariance a určuje míru lineární závislosti zkoumaných veličin. Často je také zaváděn pojem korelačního koeficientu nacházejícího se v intervalu hodnot od -1 do 1 . Koeficient blíží se 0 značí lineární nekorelovanost a díky tomu lineární nezávislost. Základní korelační koeficient bývá reprezentován jako

$$\rho_{xy} = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2 \sigma_y^2}}, \quad (3.3.3.1)$$

značící poměr kovariance veličin X, Y ku odmocnině součinu jejich rozptylů σ_x, σ_y . Vzniká opět symetrická matice $\boldsymbol{\rho}$ (3.3.3.2), tentokrát vzájemných korelací s hodnotami 1 na diagonále a párovými korelacemi mimo hlavní diagonálu.

$$\boldsymbol{\rho} = \begin{pmatrix} 1 & \rho_{xy} \\ \rho_{yx} & 1 \end{pmatrix}. \quad (3.3.3.2)$$

3.4 Barevná segmentace pomocí K-means shlukování

Za poslední desetiletí získala tato matematická metoda největší popularitu zejména v oboru strojového učení [36, 37], ačkoliv prvně ji uvedl S. Lloyd již v roce 1982 [38]. Dnes je především známa jako „unsupervised learning algorithm“ (algoritmus učení bez učitele). Tento název vyjadřuje, že pro rozdělení (segmentaci) dat do daných tříd nejsou nutná referenční (trénovací) data, na kterých by se mohla tato metoda učit. Pro oddělení jednotlivých tříd je tedy nutné využít nějakého druhu metriky, kterou zde zastává *euklidovská vzdálenost* (popsána 3.4.0.1) jednotlivých datových bodů. Z toho vyplývá, že o pravděpodobnostním rozdělení vstupních dat není nutné mít apriorní informaci. Kritickým parametrem je zde ale počet tříd K , do kterých musíme data rozdělit, abychom mohli dosáhnout požadovaných výsledků. Na rozdíl však od pravděpodobnostního přiřazování bodů použitého například u klasifikace pomocí GMM (Gaussian Mixture Model) [39] jsou body tříděny takzvaně natvrdo a patří tak pouze k jedné z K tříd.

K-means algoritmus je dopodrobna popsán v [1]. Zde uvedeme pouze základní variantu využívanou v naprogramované klasifikaci oblaků v kapitole 6.

Jak bylo popsáno výše, hlavní metrikou pro oddělení jednotlivých tříd a bodů od sebe bývá použita *euklidovská vzdálenost* \mathcal{D} (existují však i jiné). Ta vychází z *euklidovské normy* a můžeme ji pro dva body \mathbf{x}, \mathbf{y} v n dimenzionálním prostoru uvést jako

$$\mathcal{D}(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{k=1}^n (\mathbf{x}_k - \mathbf{y}_k)^2}. \quad (3.4.0.1)$$

Obecně k-means segmentace cílí na minimalizaci chybové funkce, v tomto případě již zmíněné metriky vzdálenosti od zvolených středů $\mu_K = \{\mu_1, \dots, \mu_K\}$ jednotlivých

K tříd [38]. V momentě, kdy je každý bod přiřazen ke třídě, jsou středy přepočítány podle průměrných hodnot všech bodů jednotlivých tříd a přesunuty na takto zjištěné nové pozice. Následně je znovu minimalizována metrika a upraveno přiřazování bodů. Takto k-means algoritmus konverguje do pozice, kdy již nejsou žádné nové body přiřazovány a třídy se nijak nemění. Výše nastíněným postupem minimalizujeme chybovou funkci J a řešíme takzvanou optimalizační úlohu [40]. Tu lze vyjádřit následujícím vztahem

$$\arg \min J = \sum_{i=1}^K \sum_{j=1}^{n_i} \|x_j^i - \mu_i\|^2, \quad (3.4.0.2)$$

x_j^i - výchozí data (reprezentovány body),

μ_i - přiřazené středy shluků K tříd,

K - počet tříd,

n_i - množství vstupních dat,

$\|x_j - \mu_i\|$ - euklidovská vzdálenost mezi bodem ze souboru dat a středem shluků.

Takto navržený algoritmus byl v minulosti několikrát vylepšen a dnes můžeme nalézt značné množství různých postupů, jak tuto úlohu řešit [41–45]. Za zmínění určitě stojí úprava na takzvaný k-means++ algoritmus [46]. Zlepšení se týká především náhodné inicializace středů shluků pro maximalizaci správných výsledků segmentace, abychom mohli zajistit co nejrychlejší konvergenci. Není totiž neobvyklé, že algoritmus dokonverguje do lokálního minima chybové funkce, což ovšem není správné řešení, a je proto nutné celý postup opakovat s jinými náhodnými středy tříd.

K-means++ algoritmus, který je použit v této práci pro detekci mraků, volí náhodně pouze první střed μ_1 K tříd. Další středy vybírá ze souboru vstupních dat podle pravděpodobnosti úměrné vážené druhé mocnině vzdálenosti nejbližšího bodu náhodného středu μ_1 , dokud není přiřazeno μ_k středů. Uvedený postup podstatně zlepšuje vlastnosti k-means segmentace a umožňuje nastavení menšího počtu iterací celkového výpočtu algoritmu.

V poslední době se také díky nově zrealizovaným matematickým postupům k-means segmentace a nárůstu použití tohoto algoritmu ve strojovém učení objevily nové možnosti se změnami metrik na různá jádra vyhovující lépe zkoumaným datům, a díky tomu zlepšující výsledky například pro atypicky rozložená data [47–49]. Zmíněné metody dosahují velmi zajímavých výsledků a mohly by být předmětem zkoumání pro vylepšení algoritmu klasifikace mraků v budoucnu, jejich podrobné zkoumání však pro tuto práci není potřeba.

4 Zpracování dat z celooblohového kamerového systému WILLIAM

V původní bakalářské práci [1] bylo k hodnocení barevné reprezentace atmosferických jevů zpracováno 792 snímků ze širokouhlého systému WILLIAM. Jelikož tyto snímky pokrývaly pouze 2 měsíce z celého roku, museli jsme přistoupit při tvoření obsáhlejší databáze k výraznému navýšení jejich počtu. Tato diplomová práce pracuje s databází 2044 anotovaných snímků denní oblohy ze 62 dnů napříč celým rokem. Veškeré fotografie pro vytvořenou databázi byly získávány z pozorovací stanice v Jarošově nad Nežárkou (viz kapitola 2.1). Rozšíření databáze se také týkalo nového přidání počtu parametrů k jednotlivým snímkům, ty jsou popsány dále. Díky takto obsáhlému souboru dat bylo možné v této práci vyvíjet nové algoritmy jak pro detekci, tak pro klasifikaci oblaků a testovat tak námi navržené postupy podstatně efektivněji.

V následujících kapitolách bude rozebrána upravená tvorba těchto dat, ruční klasifikace, třídění a anotace, jednotlivé výpočty a software, který byl k tomuto zpracování nutný.

Je důležité zmínit, že postup popsany níže hodnotí a popisuje, jakým způsobem byla vytvářena ruční klasifikace jednotlivých skupin oblaků a jak byla vytvářena trénovací data. Jelikož jsou z velké části některé postupy předzpracování snímků totožné i s vytvořenou automatickou klasifikací oblaků, kterou se zabývá kapitola 6, bude často tato skutečnost v následujících řádcích zmiňována.

4.1 Předzpracování snímků

Veškeré snímky, získané ze systému WILLIAM, byly z důvodu co nejmenšího zásahu softwaru a komprese fotoaparátu pořizovány ve formátu RAW, respektive NEF typického pro modely od značky Nikon. Ty bylo následně nutné důkladně předzpracovat a zajistit, aby na nich byly prováděny jasně definované transformace a deterministické operace, které by šlo následně opakovat na nově příchozích snímcích pro automatickou klasifikaci atmosferických jevů. Původně byly obrázky konvertovány do bezeztrátového formátu TIFF, aby bylo možné snímky jednoduše načítat ve výpočetním prostředí MATLAB¹⁵ a snáze s nimi manipulovat. TIFF formát velice dobře postačoval i pro tvorbu automatické klasifikace a proto byl v aktuálním procesu zpracování dat zachován.

V původní verzi databáze všechny konverze mezi formáty a vyvážení bílé obstarával zvolený software CAPTURE NX-D¹⁶ od firmy Nikon [1]. Po dokončení anotace a zkompletování celé této nově tvořené a rozšířené databáze bylo od uvedeného programu upuštěno. Problémem bylo především nemožnost definovat procesy, které byly se snímky v tomto programu prováděny. Zvolili jsme tedy open-source program DCRAW¹⁷, který zmiňované potíže odstranil (podrobněji viz kapitola 6) a plně nahradil funkčnost konverze formátů a vyvážení bílé.

¹⁵<https://www.mathworks.com/products/matlab.html>

¹⁶<https://www.nikonimglib.com/ncnxd/>

¹⁷<https://www.cybercom.net/dcoffin/dcraw/>

Vyvážení bílé [1] jednotlivých fotografií na přednastavených 6000K pro zajištění stejného barevného podání denní oblohy bylo velice důležitým faktorem. Tuto konverzi taktéž obstaraly výše zmíněné programy.

Dalším přidaným procesem, především pro snížení výpočetní náročnosti a zároveň replikovatelnosti pro automatickou klasifikaci, bylo převzorkování obrázků na zvolených 800x1200 pixelů. Tím jsme mohli zajistit snížení dat, na kterých byla prováděna segmentace, zároveň však tento krok nenarušil průběh klasifikace, neboť barevné hodnoty zůstaly zachovány. Převzorkování obsahovalo klasický typ bilineární interpolace [1, 50], který se standardně při zpracování digitálních snímků využívá a je implementován jak v prostředí MATLAB, tak zároveň pro různé knihovny pracující s jazykem Python, ve kterém byla tvořena automatická klasifikace.

Již v [1] se z důvodu špatné segmentace a oddělování jednotlivých typů oblak od sebe přistoupilo k vytvoření barevné masky odstraňující nechtěné statické objekty ze snímané scény. Opět jsme tento postup zachovali pouze s rozšířením části masky na nové statické objekty na okrajích snímků.

Jak bylo uvedeno v předchozí kapitole, protože základní barevný prostor sRGB nevyhovoval dostatečně a především správně pro segmentaci jednotlivých skupin mraků, veškeré snímky byly převedeny na barevné prostory CIE XYZ (3.2.1) a CIELAB (3.2.2). Z nich byla odseparována jasová složka Y, respektive L, a segmentace probíhala pouze na zbylých čistě chromatických složkách x, y a a*, b*. Odstranění jasové složky bylo kritické především kvůli výsledkům následující k-means segmentace. Barevné prostory jako například prostor RGB v každé ze svých složek uchovává informaci o jasu, která barevné rozložení oblaků i nebe významně ovlivňuje a díky tomu jsou výsledky segmentace značně nepřesné. Na tento jev upozornila již předchozí analýza v [13] a byla potvrzena i v mé bakalářské práci [1].

Po dokončení předzpracování mohly být snímky použity pro samotnou k-means (3.4) segmentaci, kdy byla aplikována implementovaná funkce z prostředí MATLAB, později pro real-time detekci a automatickou klasifikaci z modulu scikit-learn¹⁸ vytvářené pro Python.

Důležité téma v [1] představovalo, jaký počet clusterů je pro segmentaci vhodný. Empiricky jsme rozhodli, že nejlepší výsledky dosahujeme pro 4 clustery, kdy vždy jeden cluster zaujímal nechtěné okraje snímku zastíněné barevnou maskou. Z uvedeného vyplývá, že samotná denní obloha byla dělena na 3 různé segmenty podle své barevné podobnosti a cluster s okraji zůstal nezapočítán.

4.2 Vytvoření databáze a anotací kamerových snímků

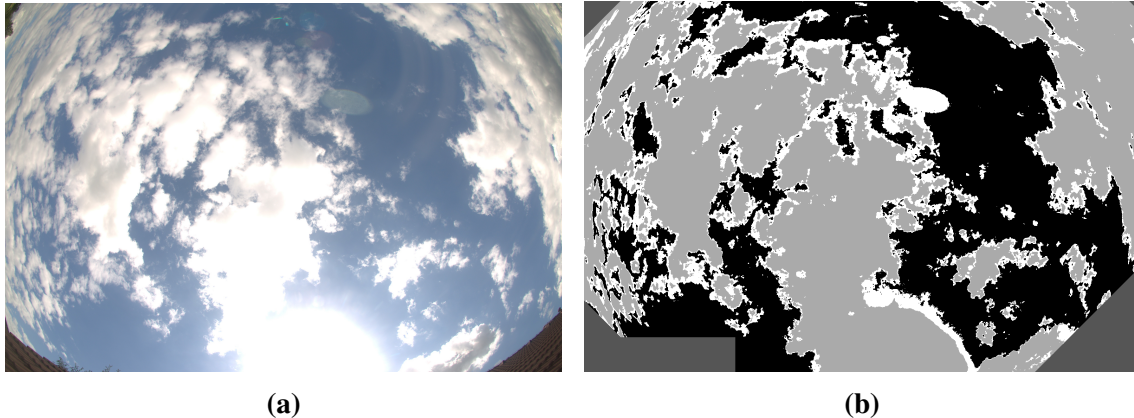
Naprosto nezbytným předpokladem pro tvorbu databáze typu mraků bylo natrénování a naučení se rozpoznávání druhů oblačnosti na denní obloze.

Ruční klasifikace spočívala v přiřazení jednotlivých segmentovaných clusterů k druhům oblaků, které se na snímku nacházely. Ty musely být co nejpřesněji subjektivně klasifikovány. Následně mohly být tyto clustery anotovány a přiřazeny do jedné z podobnostních tříd mraků (3.1.1).

Za každý prvek v databázi byl považován jednotlivý snímek, který obsahoval několik atributů (viz 3, případně příloha B).

Mezi atributy snímku patřily jejich název, číslo v celkové databázi, číslo určující pořadí pro každý den, datum a čas, kdy byl snímek vytvořen. Dále v případě, že se

¹⁸<https://scikit-learn.org/stable/>



Obrázek 4: Obrázek (a) ukazuje snímek denní oblohy, získaný ze systému WILLIAM v Jarošově nad Nežárkou. Obrázek (b) naznačuje výsledek k-means segmentace předzpracovaného snímku (a). Jednotlivé segmenty jsou následně přiřazovány ke skupinám oblaků, a správně je odstraněn nechtěný segment zakrývající statické rušivé objekty. Pro co nejlepší segmentaci výsledných oblaků či čistého nebe bylo přistoupeno k odseparování jasové složky od barevných. Proces segmentace následně probíhá pouze v chromatických komponentách a jak již bylo potvrzeno v předchozí bakalářské práci [1], výsledky jsou tím značně zpřesněny.

na nějakém clusteru nacházelo slunce, bylo nutné určit, na kterém a následně ho označit. Zaznamenáván byl taktéž cluster, který odděloval okraje snímku. Parametr stav kamery určoval, zda-li se systém nachází pokryt srážkami, či nikoliv. Nejdůležitějšími parametry byly bezesporu samostatná klasifikace oblaků, kdy jednotlivé clustery 1-4 byly subjektivně řazeny k 11 druhům oblačnosti a následně také do podobnostních skupin (3.1.1).

Zmíněný postup probíhal ve dvou verzích jak pro prostor CIE LAB tak pro CIE XYZ, čímž se podstatně zvyšovala časová náročnost, která byla nutná k vytvoření takto obsáhlé databáze.

Jelikož přiřazování přesných druhů oblačnosti je úkol značně náročný a není možné určit jeho dokonalou spolehlivost, museli jsme dále pracovat pouze s podobnostními skupinami, kde třídění probíhalo s mnohem větší přesností a mraky byly řazeny mezi další se stejnými vlastnostmi.

V této diplomové práci jsme se rozhodli do naší databáze přidat atributy, které mohly mít vliv na barevné hodnoty jednotlivých skupin mraků v průběhu dne. Jsou to parametry založené na slunečním cyklu: výška slunce nad obzorem (elevace slunce) ve stupních, časová vzdálenost od východu a západu slunce v hodinách a časová vzdálenost od pravého poledne rovněž v hodinách. Analýzou těchto příznaků zkombinovanými s barevnými hodnotami mraků se zabývá kapitola 5. V následující podsekcí jsou pro úplnost krátce uvedeny a především je zde ukázán postup jejich výpočtu.

4.3 Solární parametry

Jelikož jsou parametry popsané níže důležitým prvkem a hojně využívané v této práci, je nezbytné, abychom tyto pojmy definovaly. Zároveň díky tomu, že se nacházíme v kapitole o zpracování dat, můžeme patřičně ukázat, jak jsme při výpočtech postupovali a tyto parametry získali. Hlavním důvodem, proč jsme tyto parametry přidali do této práce, byla měnící se intenzita světla a barevnost v průběhu dne. Chtěli jsme proto zjistit, jak

moc se tyto vlastnosti projeví a jak se bude přes den měnit barevná informace, oproštěna od jasové složky, pro jednotlivé mraky nebo čisté nebe. Museli jsme proto zvolit vhodné charakteristiky, podle kterých by bylo možné výše zmíněné jevy vyhodnotit. Dále jsou tyto solární parametry uvedeny a popsány.

4.3.1 Právě poledne

Pravým polednem nazýváme stav, kdy slunce dosahuje nejvyššího bodu na obloze a zároveň prochází lokálním meridiánem [51, 52], neboli kulminuje. Lokálním meridiánem je myšlena fiktivní nebeská kružnice, procházející přes nebeský jižní a severní pól, zenit (nejvyšší bod meridiánu pro lokální území), nadir (nejnižší bod meridiánu pro lokální území) a nejj jižnější a nejsevernější bod horizontu. Z této definice vyplývá, že meridián je k horizontu kolmý. Právě poledne tak může být hodnoceno jako důležitý časový parametr nastávající v různou denní dobu, závisující na zeměpisné délce (meridiánu) a datu daného roku pro zvolenou lokaci. Díky tomu se často podstatně liší od určeného poledne ve 12.00 lokálního časového pásma.

Rozhodli jsme se tedy určit pomocí jednoduchého výpočtu

$$\Delta_{sn} = |t_p - t_{sn}| \quad (4.3.1.1)$$

časovou vzdálenost Δ_{sn} od doby pravého poledne t_{sn} pro každý jeden z pořízených snímků v určitý čas t_p a tuto závislost použít jako parametr pro další výzkum změny barvy mraků.

Časy pravého poledne jsme získali z internetového zdroje¹⁹, kde pro specifickou lokaci v Jarošově nad Nežárkou (2.1) bylo možné exportovat tabulku znázorňující jeden celý rok. Tyto hodnoty jsme dále využili v prostředí MATLAB a dávkově spočetli parametry časové vzdálenosti od pravého poledne uložené ve vytvořené databázi.

4.3.2 Elevace slunce

Elevace slunce, uváděná taktéž jako elevační úhel slunce ϕ_e , doplňkový k zenitovému úhlu ϕ_z , značí nadmořskou výšku slunce a tedy jeho polohu na obloze. Technicky také znázorňuje úhel mezi horizontem a středem slunečního disku během denního cyklu. Pomocí vzorce [53] lze vypočítat jako

$$\phi_e = \arcsin [\sin \delta \cdot \sin \psi + \cos \delta \cdot \cos \psi \cdot \cos h]. \quad (4.3.2.1)$$

Jednotlivé parametry zastupují elevaci slunce ϕ_e , deklinaci δ ²⁰, lokální zeměpisnou šířku ψ a hodinový úhel h ²¹.

Jelikož se jednalo o složitější výpočet s velkým množstvím dat z databázových snímků, využili jsme volně stažitelný předprogramovaný MATLAB skript²², který spočítal poměrně věrohodně a přesně elevaci slunce podle zadané zeměpisné délky, šířky a času pro každou z fotografií v danou hodinu. Výsledky byly ověřeny podle NOAA²³ (National Oceanic and Atmospheric Administration) kalkulátoru, věnujícímu se exaktním výpočtům sluneční pozice a dalších solárních parametrů.

¹⁹<http://www.solar-noon.com/>

²⁰<https://www.britannica.com/science/declination>

²¹<https://www.britannica.com/science/hour-angle>

²²<https://www.mathworks.com/matlabcentral/fileexchange/23051-vectorized-solar-azimuth-and-elevation-estimation>

²³<https://www.esrl.noaa.gov/gmd/grad/solcalc/>

4.3.3 Časová vzdálenost od východu / západu slunce

Z důvodu rotace země kolem své osy jsou na určitém místě nebeské sféry pozorovatelné jevy východu a západu vesmírných objektů, jako například hvězd, slunce, měsíce a dalších. K nejzásadnějším bezesporu patří východ a západ slunce, jež je běžně definovaný jako přesný čas v určitou denní dobu, kdy horní okraj slunečního disku dosáhne horizontu (definice [54]). Samozřejmostí je rozličná doba tohoto jevu pro různé lokace na zemském povrchu a roční dobu kvůli astronomickým cyklům země. K získání přesných časů východů a západů slunce jsme využili zdroj meteogram²⁴, který obsahoval tabulku těchto údajů, kterou jsme mohli exportovat do textového souboru s názvem lokace pro další možné využití při automatické klasifikaci oblaků.

Časová vzdálenost od východu a západu slunce byla poté už snadno vypočitatelná, opět například v prostředí MATLAB pomocí vzorce

$$\Delta_s = \min |t_p - t_{v/z}|, \quad (4.3.3.1)$$

kde Δ_s je časová vzdálenost, čas vytvoření snímku t_p a čas východu nebo západu slunce $t_{v/z}$ podle toho, která z těchto 2 vzdáleností byla menší. Takto vytvořené hodnoty byly opět uloženy do databáze, čímž jsme ji zkompletovali.

V tabulce 3 je pro nedostatek místa ukázán pouze zkrácený úsek popisované databáze, obsáhlejší část obsahuje příloha B.

Tabulka 3: Ukázka z finální podoby vytvořené databáze všech 2044 snímků. Jednotlivé sloupce značí číslo z celkového počtu fotografií, ID je zde zkráceno, ale obsahuje celý název obrazového souboru např. 2016_05_07_12_40_26.nef. Dalšími parametry jsou číslo pořízeného snímku pro daný den, datum a čas. cl_s zde určuje cluster z vytvořených segmentů snímku, na kterém je zobrazeno slunce. V úplné databázi (viz Příloha B) následuje 11 sloupců (druhů mraků), kde je subjektivně znázorněno, k jakému typu jednotlivé clustery patří. Parametr *edg.* označuje, který ze segmentů jsou nechtěné okraje snímku. *Rain* binárně představuje stav kamery, buď při srážkách rovný 1 či beze srážek rovný 0. Nejdůležitějšími částmi nově tvořené databáze jsou poslední parametry: přiřazení clusterů ke skupinám oblaků 3.1.1, čas od pravého poledne Δ_{sn} , respektive od západu a východu slunce Δ_s a elevace ϕ_e .

č.	ID	n	datum	čas	cl_s	cls. 1-11	edg.	rain	group 1-4	Δ_{sn}	Δ_s	ϕ_e
...
806	-.nef	1	7.5.16	12:40	3	CS, CUM	2	0	0-1,3-0-4	0.7	7.2	56.0
807	-.nef	2	7.5.16	12:50	1	CS, CUM	3	0	0-1,4-0-2	0.9	7.4	55.3
808	-.nef	3	7.5.16	13:00	3	CS, CUM	1	0	0-1,3-0-4	1.0	7.4	54.5
809	-.nef	4	7.5.16	13:10	3	CS, CUM	2	0	0-3,4-0-1	1.2	7.2	53.6
810	-.nef	5	7.5.16	13:20	3	CS, CUM	2	0	0-1,3-0-4	1.4	7.1	52.7
...

4.4 Selekcce použitelných dat

Po vytvoření všech parametrů k jednotlivým snímkům a dokončení jejich anotace a přiřazování segmentovaných clusterů jsme z důvodu kompatibility s prostředím MATLAB

²⁴<https://www.meteogram.cz/vychod-zapad-slunce/>

přistoupili k exportování vyselektovaných důležitých dat vhodných pro další matematickou analýzu. Tato data reprezentuje tabulka 4. Její podoba je shodná s již vytvořenou v předchozí bakalářské práci [1]. Kvůli jednoduššímu zpracování byla exportovaná data v textových souborech rozdělena po dnech podle toho, kdy byly snímky pořizovány. Každému zpracovanému dni tedy náležely 2 různé textové soubory exportované z databáze pro barevné prostory CIE XYZ a CIELAB. Veškerá exportovaná data jsou dodatečně k nalezení na příloženém CD v příloze A.

Tabulka 4: Příklad dat, která byla exportována z vytvořené databáze oblaků pro kompatibilní použití v prostředí MATLAB. Jednotlivé zpracované dny byly takto uloženy do samostatných textových souborů. První sloupec značí číslo snímku pro daný den, následuje datum a čas a označení jednotlivých clusterů na snímku přiřazené k jedné ze čtyř skupin mraků (3.1.1). Čísla zařazená do těchto skupin reprezentují již zmíněné 4 cluster snímku, na které byl rozdělen. Jedno číslo je vždy vyřazeno, protože tento cluster obsahoval nechtěné statické okraje.

č.	datum	čas	sk. mraků 1	sk. mraků 2	sk. mraků 3	sk. mraků 4
...
3	18.8.2016	15:20		1		2,4
4	18.8.2016	15:30	3	1		4
5	18.8.2016	15:40	3	1		4
6	18.8.2016	15:50	3	1		4
7	18.8.2016	16:00	4	1		3
8	18.8.2016	16:10	4	3		1
9	18.8.2016	16:20	4	3		1
10	18.8.2016	16:30	3	1		4
...

Takto připravené soubory jsme využili k získání barevných hodnot jednotlivých clusterů. Při vybrání segmentované oblasti z původního snímku byla spočtena střední hodnota (viz 3.3.1.1) všech pixelů společně se směrodatnou odchylkou (viz 3.3.1.4) nacházející se v daném segmentu a to pro všechny komponenty barevného prostoru CIE LAB, tak i pro CIE XYZ. Mimo jiné musely být odstraněny veškeré nulové hodnoty, které se ve snímku nacházely, aby nedošlo k vychýlení spočtených středních hodnot a směrodatných odchylek pro jednotlivé barevné složky segmentů. Po těchto výpočtech byly automaticky vytvořeny obsáhlé textové soubory obsahující výše popisované hodnoty. Ukázky z těchto dat společně s podrobným popisem jsou k nalezení v tabulce 5 a 6.

Tabulka 5: Ukázka barevných hodnot z vybraných dnů, které byly zpracovány v databázi mraků. Tabulka znázorňuje, o jaký den z databáze se jedná, pro jaké číslo snímku daného dne jsou přiřazeny následující hodnoty a skupiny oblaků (3.1.1), kam daný snímek, respektive jeho segmentovaná část patří. Dále se v tabulce nacházejí střední hodnoty bezrozměrných barevných složek \bar{L} , \bar{a}^* , \bar{b}^* , dohromady s jejich standardními odchylkami σ_L , σ_{a^*} , σ_{b^*} . Poslední tři pozice zaujímají již uváděné parametry: časový úsek od pravého poledne Δ_{sn} v hodinách (4.3.1), časová vzdálenost od východu či západu slunce Δ_s v hodinách (4.3.3) a elevace slunce ϕ_e ve stupních (4.3.2). Rozšířená ukázka těchto dat je k nalezení v příloze B.

den	č. fot.	skupina	\bar{L}	σ_L	\bar{a}^*	σ_{a^*}	\bar{b}^*	σ_{b^*}	Δ_{sn}	Δ_s	ϕ_e
[-]	[-]	[-]	[-]	[-]	[-]	[-]	[-]	[-]	[h]	[°]	[h]
...
22	19	4.	40.8	6.8	-1.8	0.9	-12.5	2.0	3.6	5.0	52.1
23	2	2.	57.7	14.9	-0.9	1.1	-0.6	3.1	1.7	4.9	54.3
24	13	3.	68.1	0.1	-0.3	0.8	1.8	2.0	2.8	3.9	9.5
25	4	1.	72.8	17.6	-4.3	1.7	-5.9	3.6	4.4	2.3	31.9
...

Tabulka 6: Stejně jako v tabulce 5 zde pro názornost uvádíme příklad získaných hodnot barevných indexů prostoru CIE XYZ jednotlivých clusterů daného snímku, společně se směrodatnými odchylkami. Taktéž jsou zde znázorněny údaje, o jaký den z celkové databáze se jedná, pro jaké číslo snímku daného dne jsou přiřazeny následující hodnoty, skupiny oblaků (3.1.1), kam daný snímek, respektive jeho segmentovaná část patří. Poslední tři pozice zaujímají rovněž časový úsek od pravého poledne Δ_{sn} v hodinách (4.3.1), elevaci slunce ϕ_e znázorněnou ve stupních (4.3.2) a časová vzdálenost od východu či západu slunce Δ_s rovněž v hodinách (4.3.3). Rozšířená ukázka těchto dat je k nalezení v příloze B.

den	č. fot.	skupina	\bar{X}	σ_X	\bar{Y}	σ_Y	\bar{Z}	σ_Z	Δ_{sn}	Δ_s	ϕ_e
[-]	[-]	[-]	[-]	[-]	[-]	[-]	[-]	[-]	[h]	[°]	[h]
...
51	44	4.	0.26	0.05	0.26	0.05	0.47	0.07	2.41	5.98	44.28
52	7	2.	0.25	0.02	0.26	0.02	0.48	0.03	3.94	4.18	28.07
54	9	3.	0.30	0.03	0.32	0.03	0.37	0.04	4.36	3.21	20.47
57	3	1.	0.27	0.08	0.28	0.09	0.45	0.14	5.22	2.20	10.07
...

Oproti původní bakalářské práci [1] kromě objemu dat (více jak 2500 nových hodnot) přibýly v datových textových souborech i nové zmiňované parametry elevace slunce, času od pravého poledne a času od východu / západu slunce pro každý jednotlivý snímek. Získali jsme tak velké množství nových zdrojových dat vhodných pro další analýzu. Tou se zabývá následující kapitola 5.

5 Analýza meteorologických jevů

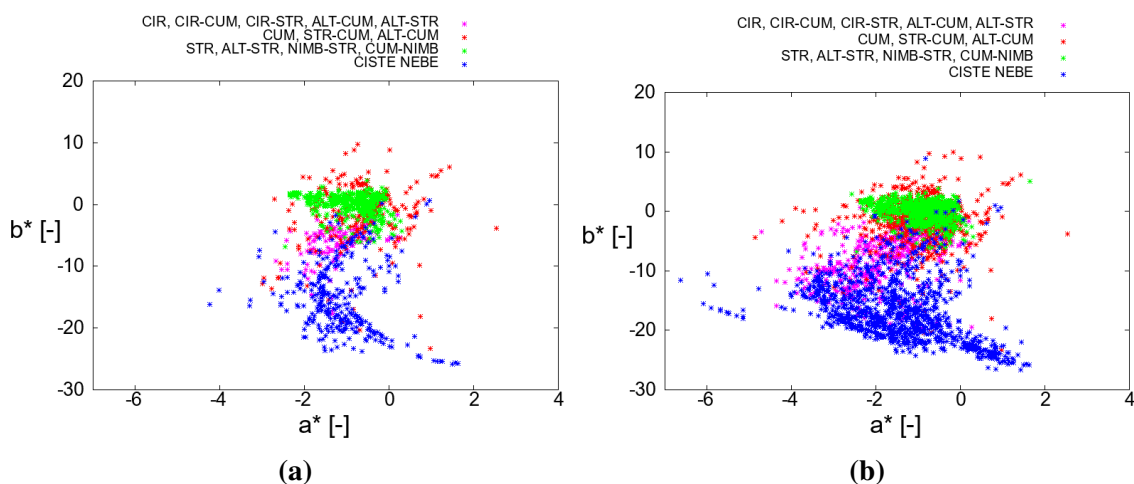
Po finální kompletaci a získání potřebných dat, která byla uložena do vytvořené databáze, jsme mohli přistoupit k jejich analýze. Prvním bodem bylo porovnání výsledků předchozí bakalářské práce [1] s mnohem obsáhlejší databází. Další body této kapitoly se věnují zobrazení jednotlivých závislostí nových parametrů a jejich evaluaci. V neposlední řadě jsme na zpracovaná data kvůli hledání nových závěrů použili různé statistické postupy snažící se odhalit další souvislosti v souboru našich parametrů. Získané informace z této kapitoly bylo následně možné aplikovat na automatickou detekci a klasifikaci oblaků popisované v kapitole 6.

5.1 Barevné diagramy mraků

V této kapitole budou ukázány naše výsledky závislostí složek barevného prostoru CIELAB, respektive CIE XYZ společně se závislostmi těchto složek na nově zvolených solárních parametrech (podkapitola 4.3). Zároveň zde představíme porovnání výsledků předchozí bakalářské práce [1] s rozšířenými daty v této práci diplomové. V závěru této kapitoly bude také navrženo, jak by se se kterými závislostmi dalo pracovat dále, a které kvůli svým nevyhovujícím parametrů byly pro další analýzu nevhodné.

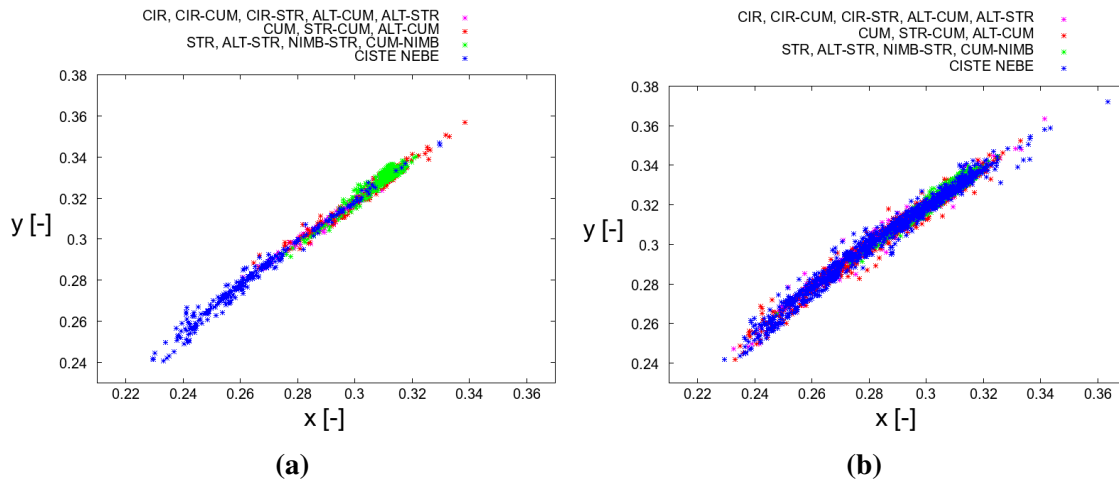
Každý ze zobrazených bodů ve všech představených grafech této kapitoly reprezentuje skupinu oblaků (3.1.1) nacházejících se na jednom snímku.

Při pohledu na diagramy na obrázku 5 můžeme vidět, že zvýšení objemu vstupních dat vytváří stejně rozložené závislosti skupin oblaků (3.1.1) v chromatickém prostoru a^* , b^* jako data z bakalářské práce [1].



Obrázek 5: Porovnání barevných hodnot a^* , b^* skupin oblaků v prostoru CIELAB pro výsledky z práce [1] (a) a obsáhlejší datový soubor získaný rozšířením databáze snímků a jejich anotací v této diplomové práci (b). Pro větší objem dat můžeme vidět, že jednotlivé skupiny jsou shlukovány do stejných prostorů v chromatickém grafu, díky čemuž lze potvrdit předchozí závěry a konstatovat, že jednotlivé skupiny lze v těchto dimenzích od sebe odlišit.

Potíže se ovšem projeví u barevného prostoru CIE XYZ, a jak již bylo naznačeno v [1], tento prostor se ukázal jako méně vhodný. Uvedený závěr je možné demonstrovat na obrázku 6, kde je vidět, že se jednotlivé skupiny oblaků se zvýšeným počtem dat ztlačují a že je již nelze dostatečně separovat. To je problém především u podobnostní skupiny mraků číslo 3 (3.1.1), obsahující dešťovou oblačnost, kterou by tak v tomto prostoru bylo těžké klasifikovat. Pro budoucí analýzu jsme tedy od tohoto barevného prostoru ustoupili a zůstali pouze u ztlačení vhodnějšího a nezávislého CIELAB.

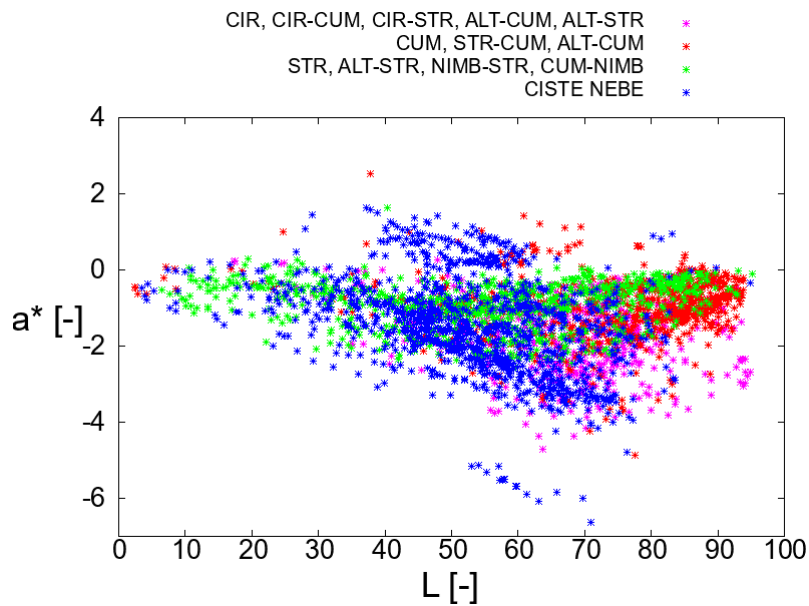


Obrázek 6: Ukázka porovnání výsledků v prostoru CIE XYZ. Obrázek (a) značí barevné hodnoty pro osy x , y z [1]. Při nižším objemu dat bylo možné pozorovat částečné shlukování jednotlivých skupin oblaků v určitých oblastech tohoto prostoru. Již v závěrech bakalářské práce [1] bylo uvedeno, že tento prostor nebyl tak vhodný pro reprezentaci barevných hodnot jednotlivých skupin mraků jako jiné nezávislé barevné prostory. Na obrázku (b) s několikanásobně větším objemem dat vidíme potvrzení této teze. V uvedeném prostoru CIE XYZ byly vykresleny hodnoty oblaků rozšířené databáze (b), na kterých je patrné, že zde již skupiny oblaků není možné dostatečně separovat. Veškeré body skupin 1-3 z bakalářské práce [1] i ty z rozšíření v této diplomové jsou zakryty skupinou oblak 4 (modrá) znázorňující čisté nebe. Naším hlavním záměrem při detekci mraků však bylo se od této skupiny co nejvíce odlišit, což zde nebylo možné. Prostor CIE XYZ je kvůli tomuto výsledku z další analýzy vynechán.

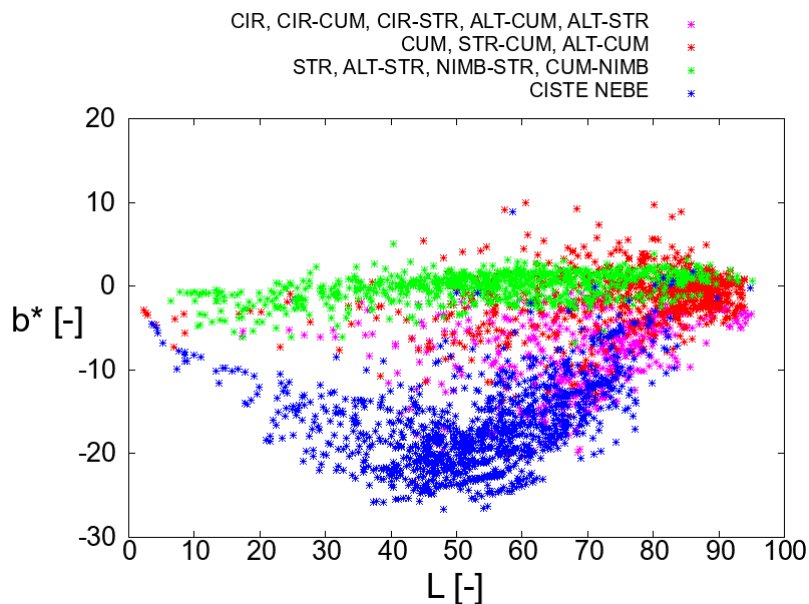
Dále jsme se rozhodli otestovat, zda by nebylo možné využít i jasovou složku, kterou jsme ve vstupních datech měli předzpracovanou a připravenou, ovšem v [1] nebyla její jakákoliv závislost zkoumána a hlavní prostor byl věnován složkám chromatickým. Vykreslili jsme tedy grafy jasové komponenty ku jednotlivým barevným složkám na obrázcích 7, 8.

Na diagramu 7 komponent L a a^* vidíme, že při změnách jasu se v chromatické složce a^* všechny skupiny oblaků vzájemně prolínají, především však skupina čistého nebe (modrá) a srážkových mraků (zelená). Toto zjištění jsme vyhodnotili pro budoucí analýzu jako problematické. Rozhodli jsem se tedy z velké části chromatickou složku a^* z dalšího výzkumu vyřadit, avšak pro úplnost zde budou ke konci sekce její závislosti na dalších parametrech uvedeny a bude demonstrována její nepřilíčná využitelnost.

Naopak graf L , b^* 8 naznačuje možnou oddělitelnost jednotlivých skupin oblaků v průběhu různých jasových hodnot. Nejlepší separovatelnost dešťové oblačnosti od čistého nebe je především v intervalu jasu od 20 do 70. Zbylé skupiny jsou v této závislosti taktéž poměrně dobře vymezené a odpovídají dalším z předpokladů v [1], kdy oblaka



Obrázek 7: Vykreslení hodnot skupin oblaků v prostoru CIELAB pro komponenty L a a^* . Při takto dané závislosti je pozorovatelné, že jednotlivé skupiny zde nejsou dostatečně separovatelné. Největší rozptyl jednoznačně vykazuje skupina číslo 4 (značená modře) obsahující hodnoty k-means segmentů pro čisté nebe. Tato skupina zde překrývá všechny ostatní, což je v kontradikci s naším cílem detekovat jednotlivá oblaka, respektive je v tomto prostoru separovat. Z uvedených důvodů nebyl tento graf pro další klasifikaci mraků využíván.



Obrázek 8: Graf komponent L a b^* znázorňující jednotlivé hodnoty skupin oblaků. Podle rozložení všech skupin je patrné, že v závislosti těchto složek bude možné od sebe oddělovat jednotlivé typy oblaků, což je přesně výsledek, který požadujeme. Nastíněná závislost byla dále využívána v automatické detekci a klasifikaci oblaků v kapitole 6.

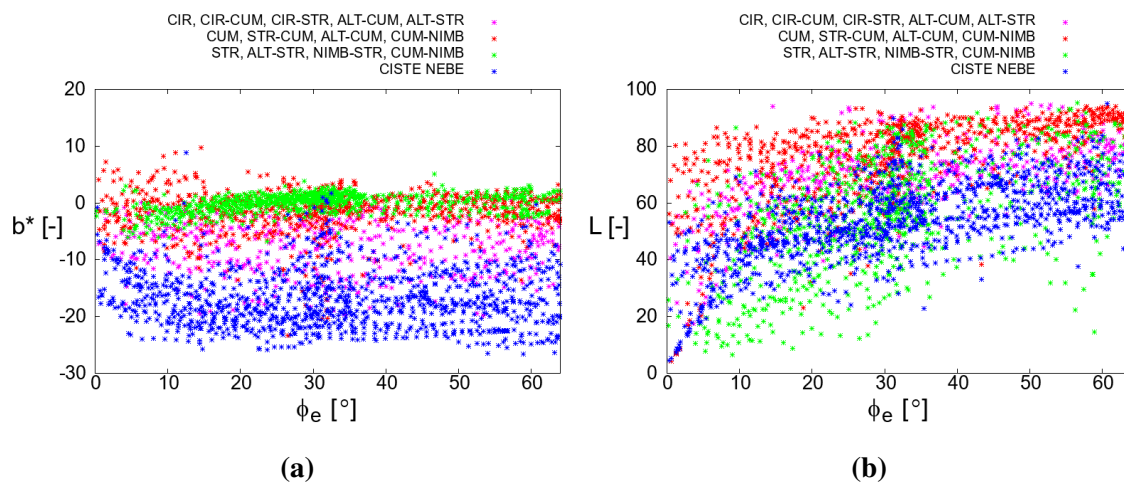
skupiny 2 (značená v grafech červeně) jsou velice podobná oblakům bouřkovým a zároveň se skupina 1 (fialová) obsahující vysokou oblačnost, blíží barevně čistému nebi. Pro nízké jasové hodnoty nebylo možné jednotlivá oblaka správně oddělit, čemuž taktéž odpovídají i závěry z [1], kde byl zmíněn fakt, že oblaka se po západu slunce nepřesně segmentovala a tvořila chybné shluky. Tento závěr bude postupně dokazován i v dalších grafech.

Závislosti L , b^* a a^* , b^* jsme díky zjištěným vlastnostem využili v automatické detekci a klasifikaci mraků v kapitole 6.

Dále se budeme zabývat nově získanými parametry, které byly přidány do databáze mraků (viz kapitola 4), a z nich tvořenými diagramy.

Započali jsme závislostmi mezi změnami komponent L a b^* v průběhu vývoje elevace slunce 4.3.2. Ty jsou znázorněny na diagramech na obrázku 9. Na prvním z nich můžeme pozorovat, že jsou skupiny mraků od sebe oddělitelné a to v průběhu celého dne, tedy při změně elevace slunce. Především skupina 3 dešťové oblačnosti se pohybuje pouze v úzkém pásmu chromatických hodnot b^* . Stejně tak tomu je i v případě skupiny 4 čistého nebe. Rozhodli jsem se tedy zkoumat, zda-li tyto jevy nepřipomínají průběh nějaké funkce. Tomuto studiu byla věnována další podkapitola (5.2).

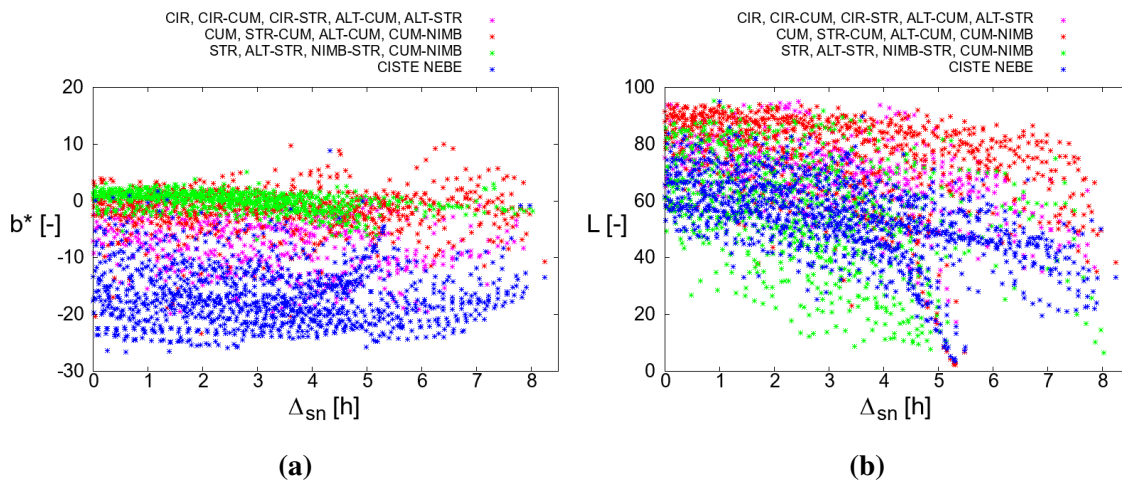
Na druhém grafu, znázorňujícím průběh jasu při změnách elevace slunce, bylo evidentní, že pro veškeré skupiny oblaků, kromě skupiny 3 (zelená), můžeme vytvořit odhad funkce, podle které se vyvíjejí. Nejvíce důležitá skupina 3 však žádnou subjektivně pozorovatelnou závislost nevykazovala a její body se nedefinovatelně prolínaly napříč všemi ostatními skupinami.



Obrázek 9: Závislosti vybraných složek prostoru CIELAB na elevaci slunce 4.3.2. Diagram (a) reprezentující barevnou složku b^* znázorňuje vývoj jednotlivých barevných skupin v průběhu celého dne, kdy se mění elevace slunce. Je patrné, že se jednotlivé skupiny přes den dostatečně separují a pravděpodobně tvoří průběhy, které budou dále zkoumány. Především skupina 3 (zelená) se srážkovými oblaky a skupina 4 (modrá) s čistým nebem jsou zřetelně rozdílné. Graf (b) se zabývá jasovou komponentou. I v tomto případě je naznačen průběh pro skupiny oblaků 1, 2 a 4. Naopak pro zbylou skupinu 3 (zelená) reprezentující srážkové mraky můžeme pozorovat neorganizovaný rozptyl dat napříč celým grafem přes všechny ostatní skupiny.

Později jsme se věnovali grafům složek L a b^* pro změnu časové vzdálenosti v hodnotách od parametru pravého poledne 4.3.1 a znázornili je na obrázku 10. Diagram se složkou b^* připomíná předchozí (obr. 9) s hodnotami elevace, pouze s převrácenou spodní

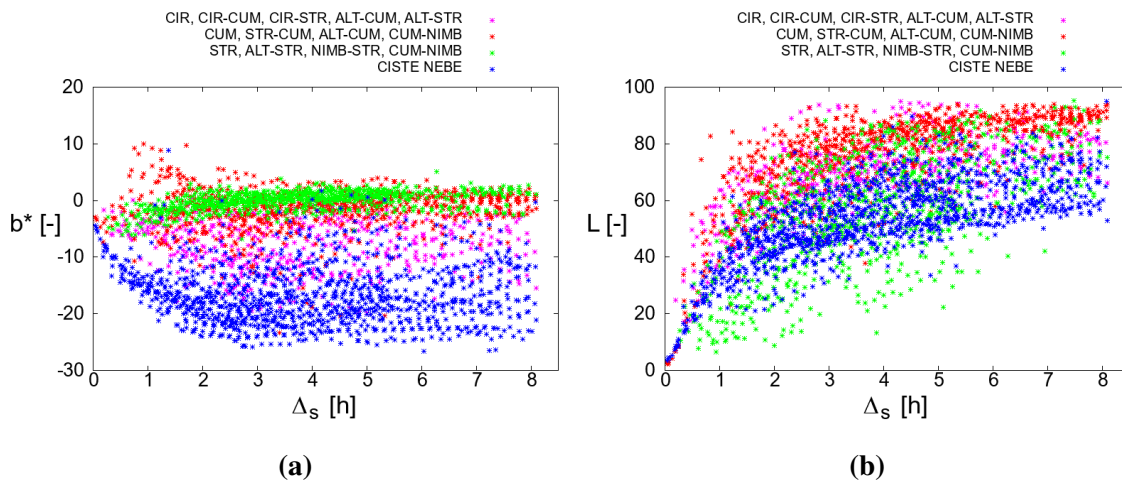
osou. Rovněž se zde odděluje skupina bouřkových a dešťových oblak se skupinou čistého nebe těmi zbylými. Veliký rozdíl je patrný na druhém diagramu věnujícímu se jasů v závislosti na času od pravého poledne. Skupiny bezesrážkové opět naznačují existenci analytické závislosti a srážková skupina 3 (označená zelenou barvou) se prolíná mezi a je velice obtížné ji od ostatních separovat. Projevil se zde však faktor různého ročního období, během kterých byly snímky pořizovány. Jelikož svítání a stmívání má rozdílné doby v průběhu roku, mění se i časová vzdálenost od pravého poledne, která byla značně viditelná na tomto diagramu a odpovídala i poklesu jasových složek v jiných hodnotách této vzdálenosti. Tento jev nám činil potíže především při hledání možných průběhů jednotlivých skupin. Nejednalo se však o nijak závažný problém a do budoucí analýzy jsme se rozhodli touto skutečností nezabývat, především proto, že se zmíněný jev v dalším parametru vzdálenosti času od východu / západu slunce nevyskytoval.



Obrázek 10: Diagramy reprezentující rozdíl času od pravého poledne pro vybrané složky L a b^* . Obrázek (a) se velice podobá předchozímu 9a, pouze s opačným vývojem. Skupiny oblaků se opět separují jedna od druhé a naznačují analytické průběhy. Stejný případ platí i pro graf (b), který je velice podobný předchozí závislosti 9b opět pouze s opačným průběhem. U obou grafů se ale ztlačněji projevovala rozdílnost rozložení při různých ročních obdobích, respektive při měnícím se času východu a západu slunce po celý rok.

Poslední zkoumanou závislostí, jež nám byla umožněna díky rozšířené databázi, bylo porovnání hodnot L a b^* s časovými vzdálenostmi od východu, respektive západu slunce. Na prvním grafu obrázku 11 je vidět nejzřetelnější separace chromatických hodnot b^* mezi jednotlivými skupinami. Zároveň je zde možné nejzřetelněji zachytit známky určitých analytických průběhů. Totéž bylo pozorovatelné u druhého grafu 10b. Díky odstranění vlivu různých dob východů a západů slunce v průběhu roku, umožněného vytvořením nejkratší časové vzdálenosti od jednoho či druhého jevu, jsme mohli závislosti tohoto parametru podrobněji prozkoumat. Je patrné, že barevné i jasové hodnoty směřovaly ze stejných počátečních hodnot napříč všemi skupinami, a tudíž byly v tuto dobu (východ / západ slunce) nejhůře oddělitelné. To odpovídá již zmíněným tezím v [1], tedy špatné segmentaci oblaků za chabých jasových podmínek. Naopak nejlepší separovatelnost byla zřejmá v průběhu dne.

I v posledním grafu 11b se potvrdila domněnka, že skupinu se srážkovou oblačností není možné predikovat v jasové dimenzi během denního cyklu a je mnohem vhodnější využít závislosti na složce chromatické b^* , ukázané na obrázcích 11a a 10a. Získané závěry jsme dále využili i v detekčním algoritmu a klasifikaci oblak v kapitole 6.



Obrázek 11: Grafy ukazující vývoj hodnot složek L a b^* v průběhu časové změny od východu či západu slunce. Jelikož se zde jednalo o minimální čas vzdálenosti od východu nebo západu, byla v těchto grafech potlačena rozdílnost vývoje barevných hodnot pro měnící se roční období. Graf (a) nejlépe ze všech uvedených obrázků reprezentuje barevné změny v průběhu celého dne a nejostřeji ohraničuje jednotlivé skupiny mraků. Znatelný vývoj hodnot jasu je zřejmý i pro každou ze skupin v grafu (b). Opět pouze skupina 3 (srážkové mraky - zelená barva) měla veškeré své jasové hodnoty rozprostřené napříč celým obrázkem a zjevně protínala každou ze skupin ostatních.

Abychom předvedli i výsledky pro poslední chromatickou složku a^* a potvrdili její nepřilíživé parametry, budou zde ukázány grafy její závislosti na získaných attributech solárního cyklu.

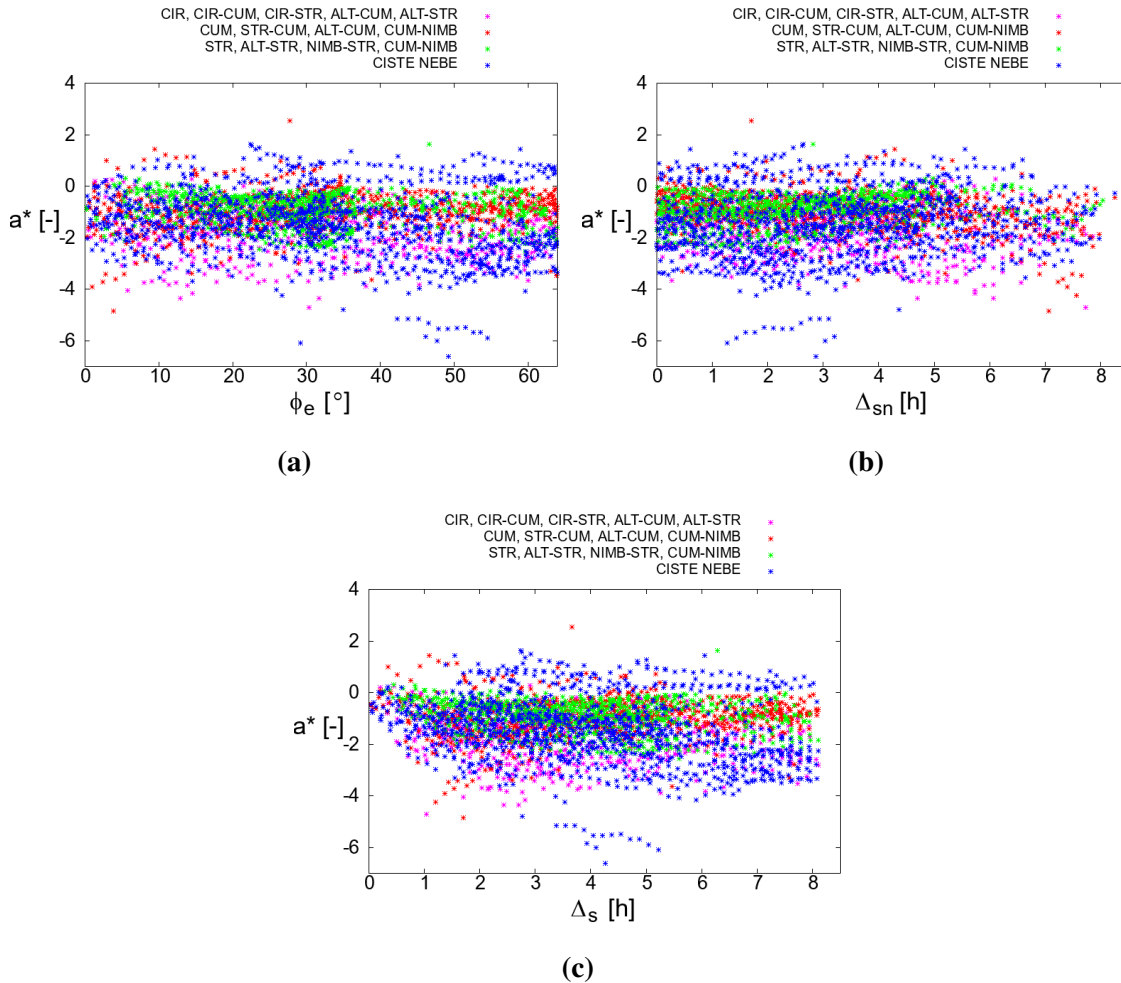
Na obrázku 12 si můžeme povšimnout téměř úplného překrytí všech skupin mraků skupinou 4 (modrá), reprezentující čisté nebe, napříč intervalu celého dne. Ani jedna závislost představených parametrů (obr. 12) nevykazuje větší využitelnost při detekci oblaků v praxi. Jediným možným rozlišením mezi skupinami jsou ohraničené intervaly v komponentě a^* , přes které se různé skupiny, především skupina 3 (3.1.1), nepřeklenou. Mimo tento interval se však jedná o malé procento dat, a toto zjištění tedy není v praxi příliš aplikovatelné. Vývoj chromatické komponenty a^* proto nebyl zkoumán ani v regresní analýze, kterou se zabývá následující podkapitola.

Dalším z parametrů, které jsme se snažili zkoumat, byla změna a posun dat v diagramu 5b na základě změny elevace, času od pravého poledne a času od východu / západu slunce. Jednalo se tak o poměrně velké množství grafů uložených do měnícího se formátu GIF (Graphics Interchange Format), ze kterých však nebylo možné vyhotovit žádné nové závěry. Data se ve zmíněném diagramu chaoticky proměňovala a nebylo možné najít alespoň přibližný vzorec jejich chování. Z toho důvodu tyto animace v této diplomové práci dále již nevyužijeme.

5.2 Analytický popis barevných diagramů

V momentě, kdy jsme měli prvotní představu o rozložení našich dat, jsme přistoupili k hledání co nejlepších analytických odhadů závislostí skupin mraků na diagramech v kapitole 5.1, neboli „fitování“. Ke zmíněné úloze nám sloužil velice vhodný Curve Fitting toolbox předinstalovaný jako součást prostředí MATLAB.

V případě hledání vývoje možných závislostí se dostáváme k úkolu regresní analýzy,



Obrázek 12: Obrázky značící vývoj chromatické složky a^* v závislosti na parametrech: (a) elevace ϕ_e , (b) času od pravého poledne Δ_{sn} , (c) času od východu / západu slunce Δ_s . Pro všechny grafy je typický již nastíněný problém na obr. 7, a to že se jednotlivé skupiny, především skupina 4 (modrá) reprezentující čisté nebe, rozptylují napříč celým grafem, bez známek možné separace. Všechny uvedené diagramy jsou si v těchto rozptylech velice podobné. Jako jeden ze závěrů však můžeme uvést, že některé skupiny, především 2 a 3, se nachází pouze v jistých ohraničených intervalech, a naopak rozptyl hodnot u zbývajících skupin je znatelně větší.

ať už lineární [34–37,55] či nelineární [55–57]. Tato úloha byla již vysvětlena v předchozí kapitole. Dále ve zkratce popíšeme, o jaký úkol se jedná, jak jsme postupovali a jaké matematické nástroje byly využity.

5.2.1 Regresní analýza

Při regresní analýze pracujeme se statistickým hledáním průběhu závislostí 2 a více proměnných. Zde se budeme zabývat pouze prostorem dvoudimenzionálním, tedy pouze se 2 proměnnými (závislými a nezávislými). Obecně hledáme funkci (regresní model), která by nejlépe aproximovala průběh dat. Snažíme se tedy najít předpis funkce, minimalizující chybu (vzdálenost) všech dat (reziduí) od této funkce, a chceme najít takzvaný nejlepší fit.

Přístupů existuje celá řada, nejvíce známá je metoda nejmenších čtverců [35–37, 55], metoda maximální věrohodnosti (Maximum-Likelihood) [36, 37], pracující s pravděpodobnostním rozložením náhodné chyby modelu, či Bayesovská regrese [36, 37]. V nejběžnějších případech dosahuje náhodná chyba modelu gaussovského rozložení [34, 35], a v tomto případě dostáváme totožné výsledky s metodou nejmenších čtverců [36, 37]. Příklady regresní analýzy budou tedy vysvětleny právě na ní. Jednotlivé případy jsou zde uvedeny pouze pro názornost a budeme se jimi zabývat v krátkosti a co nejobecněji.

Bavíme-li se o lineárním [36], respektive polynomiálním (zvláštní případ lineární regrese) regresním modelu, hledáme předpis funkce v obecném tvaru

$$t = w_0 + w_1x + w_2x^2 + \dots + w_kx^k, \quad \forall k \in N, \quad (5.2.1.1)$$

zkráceně jako

$$t = \sum_{k=0}^N w_k x^k, \quad (5.2.1.2)$$

a v tvaru maticovém

$$\mathbf{t} = \mathbf{w}^T \mathbf{X}, \quad (5.2.1.3)$$

který by minimalizoval celkovou chybu \mathcal{L} součtu druhých mocnin odchylek e vyjádřenou jako

$$\mathcal{L} = \frac{1}{N} \sum_{n=0}^N e_n^2. \quad (5.2.1.4)$$

Odchylka e_n zde reprezentuje hlavní kritérium vhodnosti modelu zápisem

$$e_n = t_n - \hat{t} \quad \text{nebo} \quad e_n = t_n - \mathbf{w}^T \mathbf{X}, \quad (5.2.1.5)$$

kde t_n značí v kartézském dvoudimenzionálním souřadném systému „ypsilonové“ hodnoty našich dat, N jejich počet, t jsou hodnoty prokládaného polynomu, \mathbf{w}^T je transponovaný vektor koeficientů polynomu, \mathbf{X} je matice „xových“ hodnot v závislosti na řádu polynomu k viz 5.2.1.6.

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^k \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_N & x_N^2 & \dots & x_N^k \end{bmatrix}. \quad (5.2.1.6)$$

Při vektorové konvenci

$$\mathbf{t}_n^T = [t_1, \dots, t_N], \quad (5.2.1.7)$$

lze v úplném maticovém tvaru, po povolených úpravách, chybu vyjádřit jako

$$\mathcal{L} = \frac{1}{N} (\mathbf{t}_n - \mathbf{X}\mathbf{w})^T (\mathbf{t}_n - \mathbf{X}\mathbf{w}). \quad (5.2.1.8)$$

Řešení minimalizace celkové chyby vede na metodu nejmenších čtverců. Zde pro minimální chybu hledáme správné koeficienty \mathbf{w} polynomu. Metodu zde odvozovat nebudeme (k nalezení zde [35–37, 55]), její výsledek pro hledané koeficienty se dá vyjádřit ve tvaru

$$\mathbf{w} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{t}_n. \quad (5.2.1.9)$$

Tímto jsou obecně nalezeny vhodné koeficienty minimalizující chybovou funkci \mathcal{L} , a díky nim jsme získali předpis pro polynom aproximující zvolená data. Záleží však samozřejmě na jeho řádu. Obecně platí, že čím vyšší polynom, tím lépe sedí zvoleným datům

(minimalizovaná chyba je menší), avšak data špatně predikuje a dochází k takzvanému over-fittingu [36]. Řád polynomu je tedy nutné volit velice pečlivě, k čemuž slouží metoda křížové validace popsaná dále (5.2.2).

Jak však bude vidět na výsledcích v sekci (5.2.4), ne vždy je možné nebo vhodné využít lineární regresi a datům lépe sedí průběh nelineární, např. exponenciální.

Používáme-li nelineární regresi, kterou prostředí MATLAB také podporuje, musíme počítat s podstatně náročnější matematickou úlohou. Řešení úlohy nelineárním modelem (odpovídá samozřejmě i lineárnímu) se hledá ve tvaru

$$\mathbf{t} = f(\mathbf{X}, \mathbf{w}) + \varepsilon, \quad (5.2.1.10)$$

kde vektor \mathbf{t} značí hodnoty závislé proměnné, f značí nelineární funkci, \mathbf{X} vektor nebo matici nezávisle proměnných, \mathbf{w} je vektor parametrů modelu, ε je náhodná variabilita odpovídající normálnímu rozložení. Řešení nastíněného problému již není zcela elementární a pro jeho zvládnutí je nutné znát celou řadu metod a algoritmů, které pro nelineární regresi byly navrženy. Můžeme uvést například metodu linearizace a jiné. Pro výsledky této práce však není nezbytné veškeré metody a jejich řešení uvádět. Odkážeme se tedy na literaturu, která se nastíněným tématem dopodrobna zabývá [55–57].

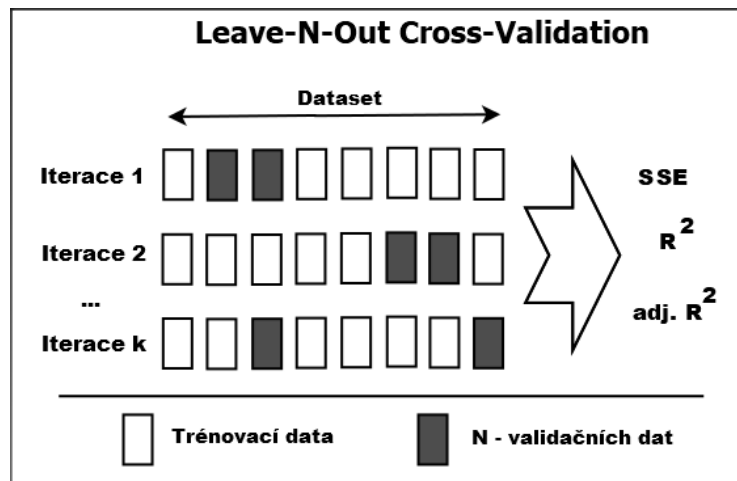
Kromě subjektivního hodnocení volby vhodnosti regresní funkce prokládané daty je důležité využít i objektivních procesů validace. Protože samotná analýza reziduální chyby modelu, jak již bylo zmíněno, není zcela dostačující, je důležité volit i metody jiné. V dnešní době se především využívají různé statistické testy, např. F-test [35], nebo také postup křížové validace (5.2.2) spojený s kritériem celkové chyby či případně s koeficientem determinace R^2 [55]. Jelikož tato metoda byla hojně využita v hledání odhadů regresních funkcí jednotlivých skupin oblaků, bude v krátkosti představena v následující podsekci.

5.2.2 Cross-validation

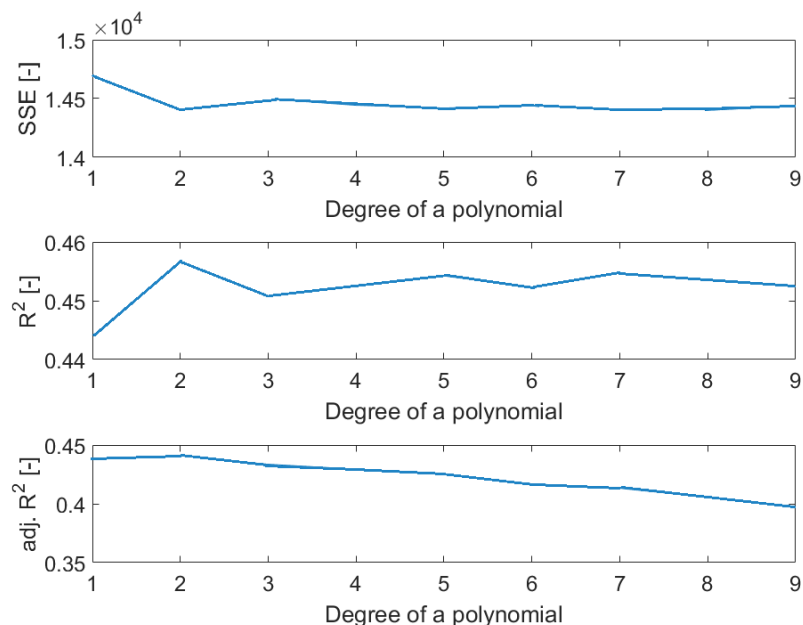
Cross-validation, česky křížová validace, je postup statistické determinace vhodného regresního modelu pro získaná data. V dnešní době existuje několik možných verzí této metody [36]. Zde bude uveden pouze ten, který jsme využili a naprogramovali pro naše účely v prostředí MATLAB. Hlavním principem je rozdělení dat na trénovací a validační skupiny. V tomto bodě se liší rozdílné přístupy k řešení takovéto úlohy. My jsme využili takzvanou Leave-N-Out cross-validaci (obr. 13), kdy z celkového souboru dat bylo vynecháno N náhodných vzorků (v našem případě 100) jako evaluační data, na kterých byla hodnocena reziduální chyba modelu, koeficient determinace R^2 [55] (validní pouze pro lineární modely) a jeho upravená verze více nezávislých proměnných *adjusted* R^2 [55, 58].

Výsledné parametry jsou zprůměrovány pro K iterací (v našem případě $K = 10000$) křížové validace a následně porovnány pro jednotlivé modely. Ačkoliv se jednalo o časově velice náročný proces, mohli jsme z těchto hodnot volit, který z analytických modelů („fitů“) nejlépe vyhovoval našim datům, a vyhodnotit, jaký z modelů zvolíme.

Příklad procesu je znázorněn na následujícím obrázku 13a a na dalším 13b je zobrazen příklad z naměřených dat.



(a)



(b)

Obrázek 13: Naznačení postupu při výpočtu křížové validace obr. (a). Náhodně je vybráno N hodnot z celkového souboru dat, které slouží jako validační, a to pro každou iteraci výpočtu regresní analýzy. Následně jsou vyhotoveny pro jednotlivé modely hodnotící parametry, které nám říkají, jak vhodný je daný model pro naše data z pohledu predikce. Na dalším obrázku (b) je znázorněn výsledek cross-validace pro rozdílné polynomiální řády regresních modelů, které byly „fitovány“ pro naše data. Naznačeny jsou zde parametry celkové reziduální chyby polynomu (5.2.3), koeficient determinance R^2 a jeho upravená verze $adj.R^2$ redukující nadměrný počet regresních parametrů. Na zvoleném příkladu je znázorněno, že celková chyba a koeficient R^2 se příliš neliší, a malý rozdíl i pro jednu nezávislou proměnnou naznačuje parametr $adj.R^2$. Tyto výsledky napomáhají určit typ regresního modelu. U koeficientu determinance platí, že čím vyšší číslo z intervalu (maximum je rovno 1), tím vhodnější je daný model pro naše data. U parametru celkové chyby (SSE) naopak hodnotíme, které hodnoty daného modelu jsou nejnižší. V tomto případě jsme podle těchto ukazatelů zvolili nejvhodnější polynomiální model řádu 2. Obecně se však při hledání nejvhodnějšího modelu jedná o velice složitou úlohu a často je potřeba dostupnosti dalších dat.

5.2.3 Predikční intervaly

Jelikož odhadované průběhy, které jsme získali v následující sekci (5.2.4), jsme chtěli využít pro klasifikaci oblaků na snímcích v reálném čase, hledali jsme pásy (intervaly) pravděpodobností, které by jednotlivým průběhům náležely a zároveň dobře predikovaly nové datové hodnoty. Díky tomu bychom tyto intervaly mohli zároveň využít pro klasifikaci k jednotlivým skupinám oblaků. Nechtěli jsme však zjistit odhad intervalu spolehlivosti [59] námi determinované regresní funkce, který by byl příliš úzký. Rozhodli jsme se tedy využít takzvaný predikční pás spolehlivosti [59], který byl pro zmíněnou úlohu vhodný a současně ho bylo možné poměrně lehce vytvořit v prostředí MATLAB.

Důležitým bodem naší analýzy bylo určení šířky predikčního pásu. Rozhodli jsme se pro standardní šířku 2σ reprezentující 95% pravděpodobnostní interval, že nová hodnota patří k dané regresní funkci. Obecně tedy predikční pás zaujímá $(1 - \alpha)$ 100% plochy přiřazující nový bod k odhadu dané regresní funkce, kde za α jsme zvolili již zmiňovaných 95%, respektive 0.05 podle tvaru předchozího výrazu.

Pro regresní funkci (5.2.1.10), hledáme pro každou hodnotu nezávislé proměnné x (označíme x_0) nový odhad hodnoty t označený \hat{t}_{new} . Předpis pro predikční interval můžeme tedy uvést jako

$$\hat{t}_{new} \pm t_{\alpha/2, (n-2)} \cdot \sigma_e \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sigma_{xx}}}, \quad (5.2.3.1)$$

kde $t_{\alpha/2, (n-2)}$ je Studentovo rozdělení [35] pro danou pravděpodobnost α s $n - 2$ stupni volnosti,

N je počet dat,

\bar{x} je průměrná hodnota nezávislé proměnné,

σ_{xx} značí rozptyl nezávislé proměnné (3.3.2.2),

σ_e je reziduální standardní odchylka spočtená jako

$$\sigma_e = \sqrt{\frac{\sum_{n=0}^N t_n - \hat{t}_n}{N - 2}}, \quad (5.2.3.2)$$

kde t_n jsou hodnoty závislé proměnné a \hat{t}_n odhady hodnot regresní funkce.

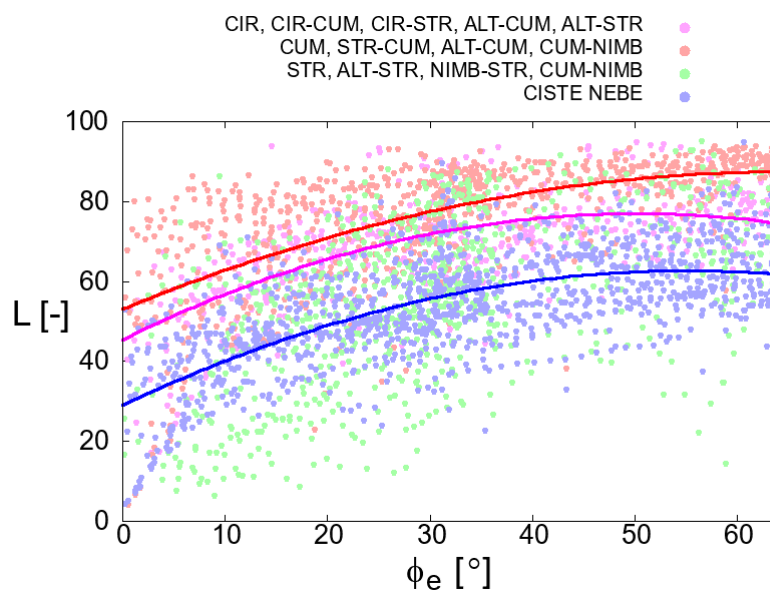
Takto jsme pro nalezené regresní křivky získali oboustranný odhad predikčního pásu, a vytvořili kritérium, do jaké skupiny oblaků by mohla spadat nově příchozí data.

5.2.4 Výsledky regresní analýzy

V momentě, kdy jsme určili, jaký postup pro hledání odhadu průběhů našich dat budeme aplikovat, a předpřipravili si veškeré funkce a skripty v prostředí MATLAB, mohli jsme přikročit k jeho realizaci. Grafy z kapitoly (5.1) zde tedy budou ukázány znovu s příklady a popisem analytických odhadů regresních modelů, které jsme vytvořili. Pro přehlednost budou nalezené funkce zvýrazněny oproti předchozím vykresleným datům, nicméně kvůli názornosti zůstanou ve stejných barvách, které jsme využívali doposud.

Mezi první grafy k hodnocení jsme zvolili ty, které se zabývají závislostmi jasové složky při změnách solárních parametrů. Prvním byl graf 14 jasu L a elevace slunce ϕ_e . Pro průběh jednotlivých skupin byl zvolen polynom 2. řádu, který podle parametrů z výsledné cross-validace (13) dával nejnižší reziduální chybu. Největší odchylka od zvolené křivky se nachází v rozmezí 0° až 10° . Zde bylo vidět, jak data ovlivňuje měnící se elevace slunce v průběhu daného roku, kvůli jinému pootočení Země a změně východů a západů

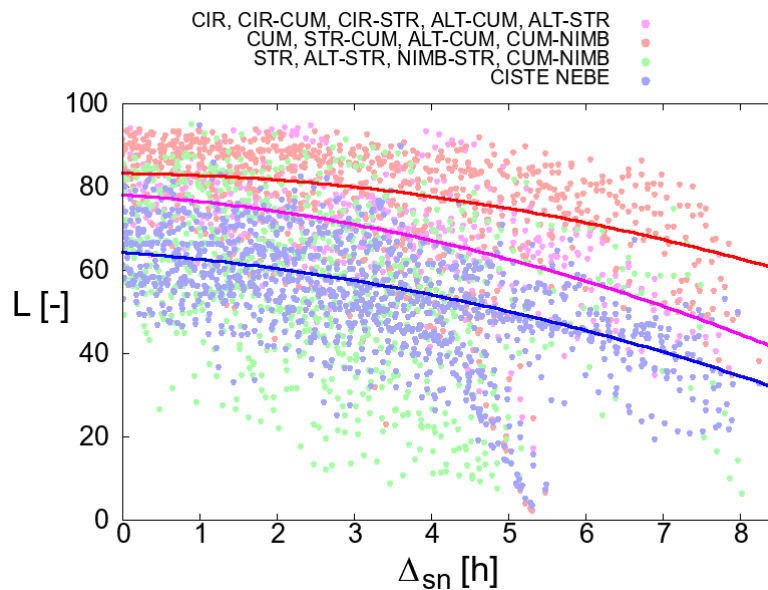
slunce. Zároveň se kvůli nízké jasové úrovni jednotlivé skupiny oblaků v tomto intervalu překrývají. Zvolená funkce nejlépe vyhovuje uvedeným datům v intervalu od 10° dále, kdy křivka poměrně věrohodně kopíruje vývoj dat jasu pro rostoucí elevaci napříč všemi skupinami. Jak již bylo uváděno při předchozí analýze této závislosti, data skupiny 3 (3.1.1) se srážkovými oblaky se rozptýlovala napříč všemi skupinami a nebylo možné vykreslené body vhodně proložit jakoukoliv křivkou, proto byl tento vývoj vynechán obecně ze všech grafů zabývajících se jasem. Na nalezených odhadech funkcí je vidět, že zbylé 3 skupiny průměrně dosahují rozdílných jasových hodnot pro vývoj elevace slunce v průběhu dne. Nejvyšších hodnot dosahuje podobnostní skupina oblaků 2 (červená) obsahující mraky typu Cumulus (3.1.1). Těm se přibližují mraky vysokého patra oblačnosti a naopak nejnižšími hodnotami jasu J jsou body reprezentující čisté nebe. Zmíněné jevy naznačují, že opravdu pro každou jednu skupinu oblaků, kromě vynechané skupiny 3 (zelená), můžeme očekávat jiné chování v průběhu dne.



Obrázek 14: Graf závislosti jasové složky L na elevaci ϕ_e . Pro 3 skupiny oblaků jsou zde ukázané odhady průběhů pomocí křivek reprezentujících polynom 2. řádu. Skupina 3 (zelená) kvůli svému rozptylu dat během změn elevace slunce neumožňovala jednoduše jakýkoliv průběh nalézt. Zbylé skupiny jsou svými odhady funkcí viditelně rozdílné a dosahují jiných maximálních hodnot během změny elevace slunce, tedy v průběhu celého dne.

Uvedenou myšlenku jiného chování podobnostních skupin oblaků nevyvrací ani další graf 15 závislosti jasu L na časové vzdálenosti od pravého poledne Δ_{sn} . Zde se, jak již bylo zmíněno z předchozí analýzy výše, nejvíce projevila změna ročního období v časech vytvoření jednotlivých snímků databáze. Jak je obecně známo, v průběhu roku se mění doba východů a západů slunce, čímž se čas od pravého poledne prodlužuje a zkracuje, a mění se tak interval, kdy mohou být denní snímky pořizovány. Tento jev nám podstatně ztěžoval co nejlepší proložení dat křivkami. V tomto případě příliš nepomohly ani objektivní metriky představené v podkapitole 5.2.2. Rozhodli jsme se tedy testovat a vybírat mezi lineárními i nelineárními modely a po zkušenosti ze závislosti elevace slunce ϕ_e a jasové komponenty L na grafu 14 jsme zvolili pro naše data polynom druhého řádu, a to pro každou podobnostní skupinu oblaků (znovu bez vynechané skupiny 3).

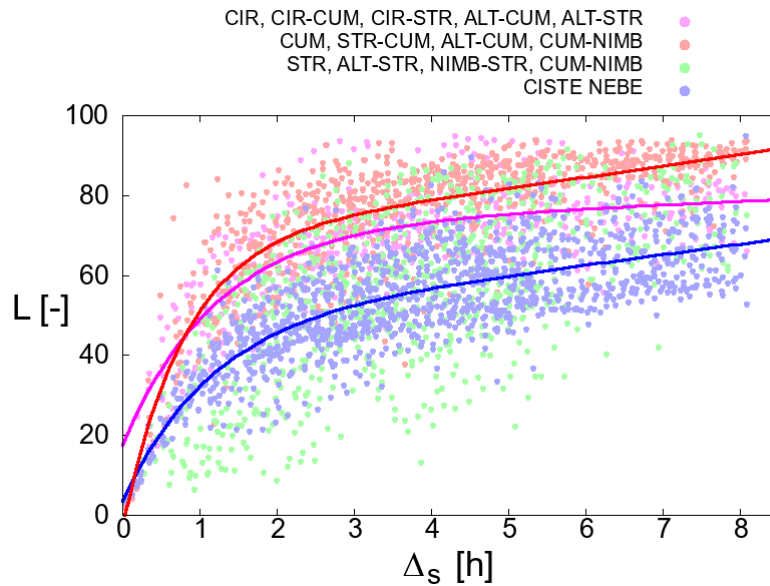
Jak naznačuje výsledný obrázek 15, i zde je možné vidět rozdílnost hodnot L průběhů jednotlivých skupin napříč celou osou Δ_{sn} , a tedy i jejich odlišení.



Obrázek 15: Graf závislosti jasu L na času od pravého poledne Δ_{sn} . Pro 3 skupiny oblaků jsou zde znovu ukázány odhady průběhů pomocí křivek reprezentujících polynom 2. řádu. Jak je možno vidět na datech v intervalu času od pravého poledne $< 5h; 8h >$, nejvíce ze všech představených závislostí se zde projevoval jev změny ročního období, respektive délky dne. Z tohoto důvodu bylo část dat z regresní analýzy vynecháno. Pro skupinu 3 (zelená) nebyl hledán odhad průběhu, opět kvůli vysokému rozptylu dat.

Z analýzy obrázku 11b můžeme vidět, že nejznatelněji pozorovatelný vývoj jasové složky L pro data skupin oblaků byl právě v závislosti na minimálním času od východu / západu slunce Δ_s . Graf 16 tyto odhady vývoje jednotlivých skupin naznačuje. Jelikož zde data nebyla zatížena nepřesnostmi kvůli měnícímu se ročnímu období, nejlépe nám podle evaluace chyby reziduí vycházely odhady exponenciálních křivek. Ačkoliv se jednalo o postup nelineární regrese (5.2), MATLAB umožňoval svými algoritmy spočítat odhady jednotlivých parametrů těchto křivek a proložit jimi tak nalezená data. Oproti předchozím případům zde odhady koeficientů exponenciál pro algoritmy nelineární regrese konvergovaly k přijatelným hodnotám a bylo možné jimi vhodně proložit naše experimentální data a to jak po stránce subjektivní, tak pomocí objektivních kritérií křížové validace. U předchozích případů (obr. 15 a 14) z důvodu divergence odhadů koeficientů pro nelineární křivky v prostředí MATLAB jsme se museli uchýlit pouze k regresi lineární. Jak si můžeme všimnout na obrázku 11b, v intervalu času od 0 až po 1 hodinu od východu či západu slunce se křivky jednotlivých skupin oblaků, kromě znovu vynechané skupiny 3 (zelená), protínaly. Tento výsledek znovu potvrzuje závěry z [1], že při nízkých hodnotách osvětlení denní oblohy, tedy těsně po východu či před západem slunce, se jednotlivá oblaka špatně segmentují, a není je možné od sebe odlišit. V dalším průběhu grafu 16 už uváděné křivky byly znatelně separovatelné a tvořily tak pro jednotlivé skupiny oblaků různý odhad vývoje jasu. Pro úplnost znovu zmíníme, že skupina 3 (zelená), stejně jako u předchozích grafů, žádnou viditelnou závislost netvořila a byla tak z analýzy vynechána.

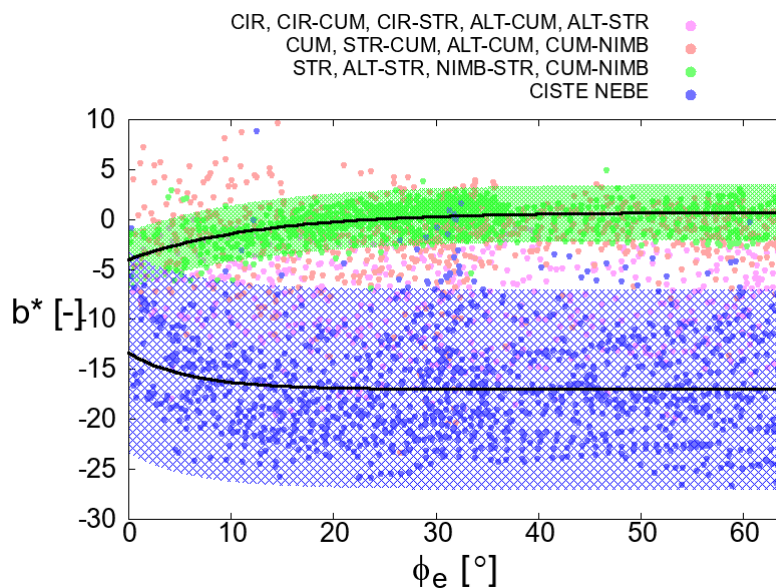
V další části jsme se zaměřili na průběhy dat pro chromatickou složku b^* a na její vývoj při změně parametrů elevace slunce, času od pravého poledne a času od východu a západu



Obrázek 16: Průběh jasové složky L v závislosti na času od východu / západu slunce Δ_s . V tomto případě datům nejlépe vyhovuje exponenciální model, a to pro všechny skupiny mraků kromě skupiny 3 (zelená). Při nízkých hodnotách jasu, respektive při krátké časové vzdálenosti východu / západu se jednotlivé odhady křivek překrývají. S narůstajícím časem se však jednotlivé skupiny opět svými odhady průběhů separují.

slunce. Již z původních grafů v kapitole (5.1) bylo zřejmé, že v barevné komponentě b^* by bylo možné podle zmiňovaných průběhů jednotlivé skupiny oblaků odseparovat. Rozhodli jsme se tuto tezi prověřit v praxi a navrhnout pro dané závislosti odhady regresních funkcí skupin oblaků. Jelikož z analýzy v podkapitole 5.1 vyplývalo, že skupiny oblaků se povětšinou pohybují jen v určitých intervalech, rozhodli jsme se k odhadům regresních funkcí přidat i odhady predikčních pásem (5.2.3), kde by se nově příchozí data mohla vyskytovat, a zároveň určit, ke které funkci by následně mohla být přiřazena. Jejich šířka byla zvolena statisticky na 2σ . Tato hodnota reprezentuje pravděpodobnost, že příchozí data spadající do daného predikčního intervalu z 95% patří k danému odhadu regresní křivky. Pro maximální přehlednost jsme regresní analýzu s predikčními intervaly prováděli na 2 nejdůležitějších a zároveň i nejlépe separovatelných podobnostních skupinách oblaků 3, 4, patřící srážkové oblačnosti (pás zelené), respektive čistému nebi (pás modré). V prvním případě (obr.17) byla zkoumána závislost b^* na elevaci slunce ϕ_e . Obě skupiny zde byly prokládány odhady exponenciálních funkcí, které nejlépe odpovídaly zadaným datům. Při pohledu na odhady funkcí obou skupin můžeme konstatovat, že skupině 3 (zelená) vyhovuje nejlépe odhad s rostoucí exponenciálou, zatímco pro čisté nebe (modrá) je přesnější klesající odhad exponenciální funkce. Mezi oběma skupinami je rozdíl i mezi šířkou odhadů jejich predikčních intervalů. Predikční interval skupiny 3 (zelená, srážková) byl podstatně užší než pro skupinu 4 (čisté nebe). Díky vyhotoveným závěrům a zkoumaným odhadům predikčních intervalů a jejich křivek, jsme našli klasifikační kritéria, která by bylo možné využít i dále pro automatickou klasifikaci.

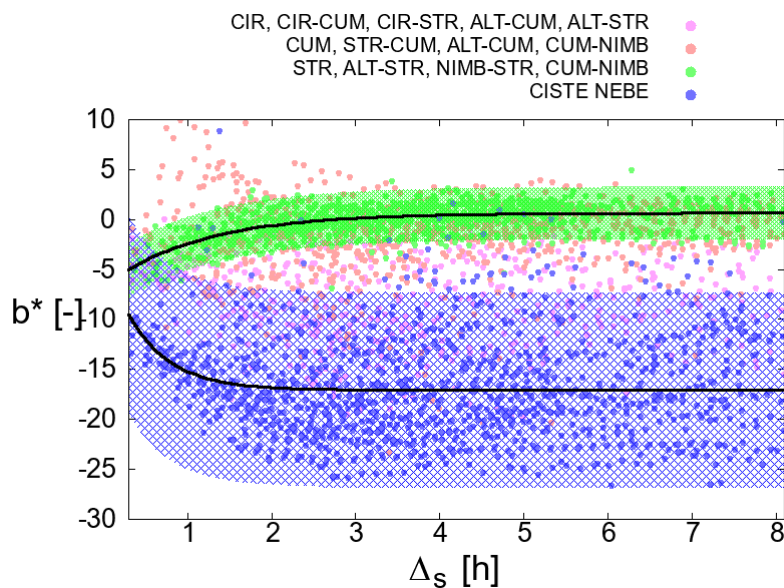
Se stejně rozloženými odhady průběhů regresních funkcí uváděných skupin 3 a 4, avšak ještě s vhodněji rozloženými daty, jsme získali graf (18) barevné komponenty b^* v závislosti na času od východu, respektive západu slunce. Oběma skupinám opět nejlépe odpovídaly odhady exponenciálních funkcí s menší reziduální odchylkou těchto odhadů



Obrázek 17: Závislost chromatické komponenty b^* na změně elevace slunce. Pro skupiny oblaků 3 (zelená, dešťové mraky) a 4 (modrá, čisté nebe) jsou zde naznačeny odhady regresních křivek a jejich predikčních intervalů. Uvedená predikční pásma jsme v dalším zpracování zvolili jako klasifikační kritérium skupin oblaků pro nově získaná data z příchozích snímků.

než pro předchozí případ elevace slunce. Obecně také můžeme hodnotit, že vypočtený odhad 95% predikčního intervalu byl užší než v předchozím grafu a tudíž se jednalo ještě o vhodnější klasifikační model.

Podle obou naposledy uvedených závislostí na obrázcích (18) a (17) lze znovu potvrdit závěry z bakalářské práce [1] zmiňující, že hodnoty skupiny oblaků 2 (červené) a 3 (zelené) se vzájemně značně překrývají v chromatické složce b^* a jsou si tedy barevně velice blízké. Tento jev zároveň potvrzuje fyzikální fakt, že většina dešťových mraků vzniká právě ze skupiny 2, označené červeně, obsahující kupovitou oblačnost.

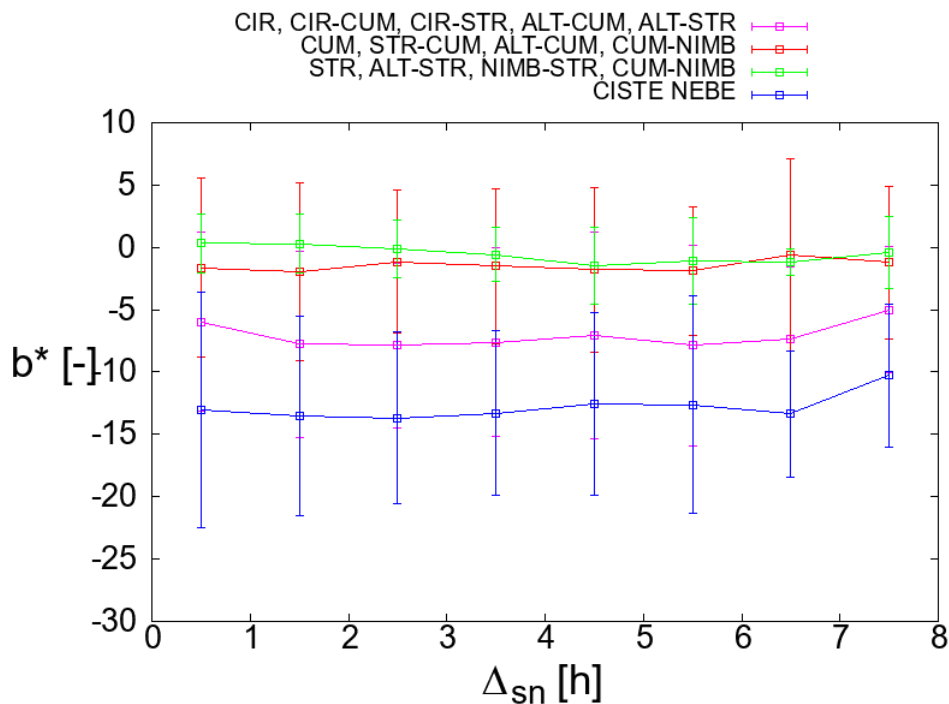


Obrázek 18: Graf chromatické komponenty b^* a změny času od východu a západu slunce. Pro skupiny oblaků 3 (zelená, dešťové mraky) a 4 (modrá, čisté nebe) jsou zde znovu naznačeny odhady regresních křivek a jejich predikčních intervalů. Uvedená predikční pásma, stejně jako v předešlém případě, byla dále využita ke klasifikaci skupin oblaků.

Rozdílný přístup jsme zvolili u posledního obrázku závislosti komponenty b^* na časové vzdálenosti od pravého poledne (obr. 19), a to především z důvodu značné podobnosti s ostatními grafy této barevné složky. Rozhodli jsme se pro interval jednotlivých hodin časové vzdálenosti od pravého poledne pro každou ze skupin oblaků vypočítat průměrné hodnoty a ty jako vývoj vynést do grafu 19. Na uvedeném obrázku je v chromatické složce b^* opět velice zřejmá podobnost bouřkových a dešťových oblaků ze skupiny 3 (zelená) s kupovitou oblačností skupiny 2 (červená). Dále můžeme pozorovat možnou separovatelnost zbylých skupin v této barevné složce, které jako v předešlých grafech v průběhu dne dosahují jen určitých hodnot. Nový přístup nám však žádné nové poznatky nepřinesl. Nemohli jsme tedy rozšířit náš model klasifikace oblaků a z tohoto důvodu uvedená závislost na obrázku 19 nebyla při dalším zpracování v následujících kapitolách využita.

Vyhodnotíme-li veškeré výsledky představené v této sekci a jejich možnost aplikovatelnosti v dále vyvíjené detekci a klasifikaci mraků, zvolili jsme pro tuto funkci jen určité a nejvhodnější modely analyzovaných dat. Těmito modely jsou závislosti z grafů 18, 17, a to především díky důvodům uvedeným ve výše představeném textu. Na těchto a dalších modelech byla vybudována automatická detekce a klasifikace oblaků a jejich vlastnosti tak budou dále reprezentovány v kapitole 6 zabývající se klasifikačním návrhem.

Protože jsme však v celé této kapitole získali mnoho parametrů pro analýzu a detekci oblaků s očividně velkým objemem redundantní informace, rozhodli jsme se otestovat redukci dat pomocí matematické metody analýzy hlavních komponent, která je představena dále.



Obrázek 19: Výpočet středních hodnot b^* bodů jednotlivých skupin oblaků společně se směrodatnou odchylkou (2σ) pro danou hodinovou vzdálenost od pravého poledne Δ_{sn} . Skupiny 2 (červená) a 3 (zelená) se zde opět protínají díky své barevné i fyzikální podobnosti. Ostatní skupiny jsou od sebe zřetelně separovatelné.

5.3 Analýza hlavních komponent

Pokud, jako v našem případě, získáme mnohazměrná data, respektive data s rozsáhlým počtem parametrů tvořící multidimenzionální prostor, je možné přistoupit k různým druhům statistických analýz pro další zpracování těchto dat. Běžně může jít o metody týkající se třídění, respektive dekorelací signálu, odstranění redundance nebo snížení dimenzionality. Rozhodli jsme se skutečnost multidimenzionality našich dat blíže prostudovat z matematického hlediska a k tomu jsme po zvážení zvolili statistickou metodu analýzy hlavních komponent.

Nejprve tuto metodu v krátkosti, bez větších detailů, představíme a v další podsekcí uvedeme získané výsledky.

5.3.1 Představení analýzy hlavních komponent

Prvně byla analýza hlavních komponent, často zkracována do anglického PCA (Principal Component Analysis), vytvořena Karlem Pearsonem roku 1901 [60]. Dnes je známá také pod názvem Karhunen–Loèveho transformace KLT [61], a její praktické využití umožnilo především nezávislé odvození Haroldem Hotellingem ve 30. letech 19. století [62].

Hlavní devízou analýzy hlavních komponent je transformace dat, která mohou obsahovat redundantní informaci a tudíž jsou korelována na data jinak reprezentována a lineárně dekorelována, neboli lineárně nezávislá.

Chceme-li formulovat předchozí tvrzení přesněji, PCA hledá malý počet lineárních kombinací korelovaných parametrů tak, aby bylo popsáno co nejvíce rozptylu v datech

co nejmenším počtem nekorelovaných parametrů [63]. Nově reprezentovaná data jsou díky tomu transformována do nového souřadnicového systému, kdy jednotlivé osy značí právě hlavní komponenty.

Odvětví, ve kterých se tato metoda používá, bývají velice různorodá, pro představu uvedeme příklady jako strojové učení, neuronové sítě, regresní analýza a jiné.

Pro výpočet se odkážeme na základy lineární algebry zabývající se vlastními vektory a čísly [64, 65]. Pro každou vstupní symetrickou matici

$$\mathbf{X}^T = \mathbf{X}, \quad (5.3.1.1)$$

můžeme tvrdit, že lze ortogonálně diagonalizovat a má pouze reálná vlastní čísla. Díky tomu existuje reprezentace ve tvaru

$$\mathbf{X}\mathbf{v}_i = \lambda_i\mathbf{v}_i, \quad (5.3.1.2)$$

kde λ_i znázorňuje reálná vlastní čísla $\lambda_1, \dots, \lambda_n$ a vektor \mathbf{v}_i nenulové reálné vlastní vektory $\mathbf{v}_1, \dots, \mathbf{v}_n$, pro n rovno počtu sloupců matice \mathbf{X} [63].

Hlavní myšlenka PCA je reprezentována rovnicí

$$\mathbf{Y} = \mathbf{X}\mathbf{V}, \quad (5.3.1.3)$$

značící transformaci původních hodnot matice \mathbf{X} pomocí transformační matice vlastních vektorů \mathbf{V} na nové výstupní hodnoty matice \mathbf{Y} , která by měla obsahovat nekorelovaná data.

Vytvoříme-li kovarianční (3.3.2.3), respektive korelační matici (3.3.3.2) \mathbf{R}_x ze vztahu $\mathbf{R}_x = \mathbf{X}^T\mathbf{X}$ (vstupní data ještě předtím centrujeme, neboli odečteme jejich střední hodnotu), můžeme následně matici \mathbf{V} získat pomocí úpravy předchozího výrazu (5.3.1.2) na

$$\mathbf{R}_x = \mathbf{V}\mathbf{D}\mathbf{V}^T, \quad (5.3.1.4)$$

kde \mathbf{V} je matice vlastních vektorů a \mathbf{D} matice vlastních čísel na diagonále.

Chceme-li rekonstruovat data s nižší dimenzionalitou, zachováme pouze vlastní vektory odpovídající počtu k nejvýznamnějších vlastních čísel (nejvýznamnějšímu rozptylu) a zbylé můžeme vynechat. Dostáváme se tedy do fáze ztrátové komprese dat, avšak při zachycení velké části variance. Jakým způsobem, respektive podle jakého klíče jsou hlavní komponenty zachovány, je možné řešit různorodě. Nejčastěji se využívá prahování [66], kdy je nezbytné určit jak velký rozptyl dat je možné zanedbat, často například 5%. V tomto případě by nám zůstala data obsahující 95% celkového rozptylu. Pro prahování se využívá vzorec

$$threshold \leq \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^N \lambda_i}, \quad (5.3.1.5)$$

kde λ_i jsou získaná vlastní čísla, N jejich celkový počet, a práh anglicky *threshold* v intervalu $< 0; 1 >$, tedy pro 95% zanechání rozptylu dat odpovídá číslu 0, 95.

Velikou nevýhodou PCA je, že se jedná o analýzu datově závislou a ne se všemi rozloženími vstupních hodnot je možné tuto analýzu provést. V takovém případě je nutné přistoupit k nějakému jinému druhu analýzy zmíněnými třeba zde [67–69] a dalším, které však v této práci popsány nebudou.

5.3.2 Výsledky PCA pro solární parametry

Pro testování PCA na našich datech jsme zvolili opět programovacího prostředí MATLAB, které představenou metodu ve svých toolboxech obsahuje a značnou výhodou je v tomto prostředí její snadná aplikovatelnost na různorodá data. Při zkoumání redundance našich solárních parametrů již napovídaly poměrně vysoké hodnoty korelace (viz 3.3.3) mezi časem od východu či západu slunce a elevací slunce. Signifikantní, avšak nižší záporná korelace byla nalezena také mezi parametrem času od pravého poledne (4.3.1) a zbylými dvěma (čas od východu a západu slunce 4.3.3 a elevace slunce 4.3.2). Vzájemná podobnost byla patrná i podle grafů 9 a 11 v předchozí podkapitole 5.1. Tento získaný předpoklad potvrdila také analýza hlavních komponent. Při aplikaci zvolené metody na všechny 3 solární parametry (viz 4.3) jsme získali jejich jinou reprezentaci ve 3 lineárně nezávislých hlavních komponentách. Při porovnání vlastních čísel přiřazeným k jednotlivým hlavním komponentám jsme zjistili, že první hlavní komponenta (označíme zde jako PC1) obsahovala více než 99% celkového rozptylu dat. Tento výsledek naznačoval, že při vyřazení zbylých dvou hlavních komponent provedeme ztrátovou kompresi ze 3 na 1 parametr reprezentovaný první hlavní komponentou při zanedbání méně než 1% celkového rozptylu dat. Rozhodli jsme tento krok učinit, zbylé dvě hlavní komponenty (označené PC2 a PC3) vyřadit, a tím snížit dimenzionalitu našich parametrů.

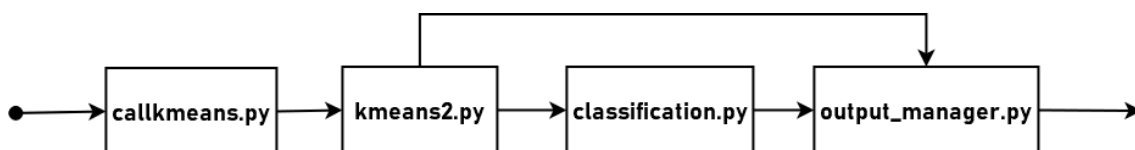
Při vykreslení první hlavní komponenty PC1 v závislosti na chromatické složce b^* , respektive L z našich dat, jsme dostali rozložení téměř totožné jako v případě závislosti b^* či L na elevaci slunce (obr. 9), ovšem v důsledku centrování, odečtení střední hodnoty (viz 3.3.1) pro jednotlivé parametry v metodě PCA, s rozdílným měřítkem osy náležící k této komponentě. Zjistili jsme tedy, že analýza hlavních komponent nám v tomto případě nepřináší další výhody použitelné jak při detekci, tak při klasifikaci oblaků, a data pomocí ní nejsou reprezentována účelnějším způsobem, kde by se od sebe jednotlivé skupiny mraků (3.1.1) vhodněji odlišovaly. Rozhodli jsme se tedy ponechat naše solární parametry v původní podobě a postavili na nich náš detekční algoritmus, který je představen v následující kapitole 6.

6 Užití algoritmu pro včasnou výstrahu při předpovědi počasí

Po provedení analýzy v kapitolách 5, 5.3, 5.1 a 5.2.4 jsme přistoupili, pro využití nově získaných údajů, k řešení praktického problému automatické detekce a klasifikace mraků. Veškerý vývoj probíhal v programovacím jazyku Python²⁵, obsahující vhodné knihovny pro řešení uvedené úlohy.

6.1 Algoritmus detekce a klasifikace oblaků

Na počátku je důležité vysvětlit veškeré prvky algoritmu pro detekci a klasifikaci oblaků a jednotlivé postupy, které zde byly naprogramovány a využity. K tomu slouží diagram na obrázku 21. Zároveň pak pro naznačení závislosti naprogramovaných skriptů v jazyce Python tvořících tento algoritmus byl pro přehled a časté odkazy vytvořen diagram 20. V tomto složení posloupnosti skriptů a návaznostech také aktuálně klasifikace mraků funguje na internetových stránkách projektu WILLIAM²⁶.

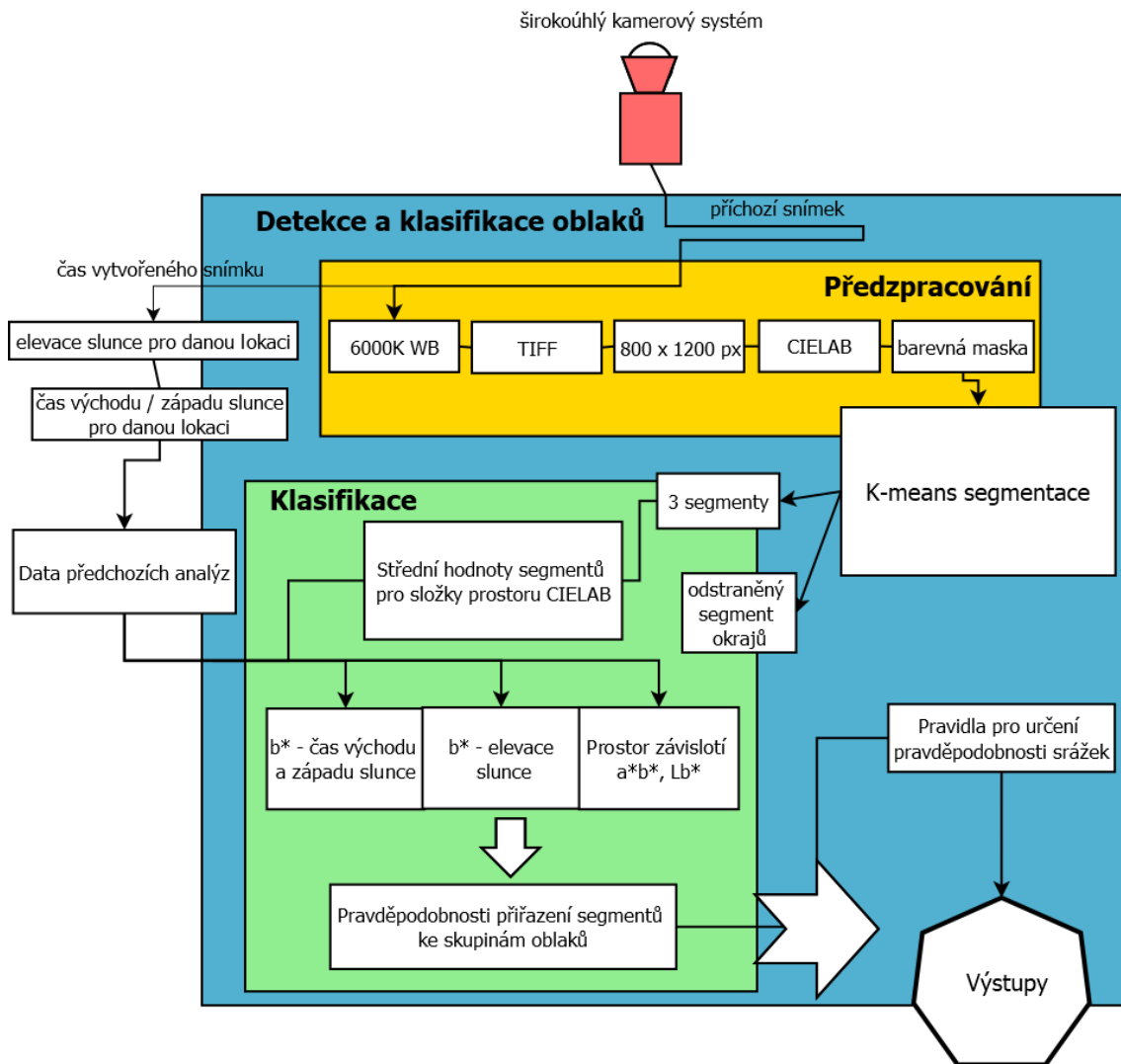


Obrázek 20: Zjednodušený diagram funkce detekce a klasifikace mraků pro jednotlivé skripty uvedené na CD v Příloze A. Názvy skriptů odpovídají odkazům v následném textu. První blok *callkmeans.py* zajišťuje schopnost volání celého algoritmu detekce z terminálu, při použití různých vstupních parametrů popsaných v dokumentaci uvnitř tohoto skriptu. Další blok *kmeans2.py* obsahuje celkové předzpracování a k-means segmentaci vstupního snímku a vytváří první výsledky segmentů tohoto snímku. S těmi následně pracuje skript *classification.py* obsahující jednotlivé metody klasifikace, načítající doplňková vstupní data a vytvářející klasifikační pravděpodobnosti společně s pravidly pro určení srážek. Poslední blok *output_manager.py* kontroluje a vytváří veškerá výstupní data z algoritmu detekce oblaků, která se dají dál využít např. v programu na tvorbu grafů Gnuplot.

Prvním bodem je získání surového snímku ze širokoúhlého systému, v tomto případě ze systému WILLIAM, ve formátu RAW a následně ho předzpracovat metodami obrazového zpracování dat. Tento postup z velké části odpovídá postupu v sekci (4.1). Snímek je pokaždé nutné připravit na k-means segmentaci. Nejdříve je na něm provedena operace vyvážení bílé na 6000K a pak je konvertován na formát TIFF probíhající v programu DCRAW. Dále je snímek (pokud je to nutné) podvzorkován na zvolených 800x1200 pixelů. Důležitým bodem je převedení na barevný prostor CIELAB. Následuje proces

²⁵<https://www.python.org/>

²⁶<http://william.multimediatech.cz/>



Obrázek 21: Diagram vytvořený a v jazyce Python naprogramovaný automatické detekce a klasifikace oblaků pro širokoúhlé kamerové systémy. Jsou zde naznačeny jednotlivé operace, které jsou při detekci prováděny, včetně předzpracování příchozích snímků, segmentace k-means, přidání dodatečných vstupních dat a konečné klasifikace s výsledky.

aplikování barevné masky odstraňující nechtěné statické objekty a to buď podle počátečních zadaných parametrů pro pozorovací stanici v Jarošově nad Nežárkou nebo v Praze. Takto upravená fotografie je odeslána na vstup algoritmu k-means++ segmentace, kde je rozdělena na 3 použitelné segmenty podle barevných podobností a 1 segment odstraňující okraje zastíněné barevnou maskou. Tento proces zastává skript s názvem *kmeans2.py*, který je možný nalézt (jako celou naprogramovanou klasifikaci) na CD v příloze A.

V momentě, kdy máme k dispozici jednotlivé segmenty, jsou pro všechny z nich spočteny střední hodnoty ve složkách L , a^* a b^* a díky tomuto procesu získáváme první viditelný výstup. Tyto střední hodnoty jsou vykreslovány do grafů závislostí L , b^* a a^* , kde máme již předpřipravené původní střední hodnoty shluků jednotlivých skupin oblaků z obrázků 5b a 8. Segmenty získaného snímku tak mohou být klasifikovány podle dopadu jejich hodnot do zmiňovaných prostorů složek CIELAB. Veškeré zmíněné výstupy budou

znázorněny v sekci 6.3.

Získání středních hodnot nám taktéž umožňuje vytvořit klasifikační metody, podle kterých budou tyto hodnoty porovnávány, respektive bude jim přidělována pravděpodobnost, ke které z podobnostních skupin oblaků patří. Tyto výpočty probíhaly ve skriptu *classification.py* (příloha A) a jsou popsány v další podkapitole 6.2.

Po vypočtení jednotlivých pravděpodobností jsou ke každému clusteru (segmentu) snímku přiřazena 4 procentuální vyjádření normovaná na 100% reprezentující pravděpodobnost, ke které ze 4 skupin oblaků (3.1.1) daný segment patří. Každá z metod klasifikace (6.2) má stejný počet výstupů, které jsou ve výsledku zprůměrovány a opět normovány na 100%. Pokaždé tak na více modelech klasifikace dostaneme celkovou procentuální pravděpodobnost výskytu dané skupiny oblaků na některém ze segmentů. Data z jednotlivých segmentů jsou následně vyhodnocována a transformována na výstupy (kapitola 6.3).

Jak bylo zmíněno hned v první kapitole, v současné době jsou v provozu 2 verze systému WILLIAM a to v Praze a v Jarošově nad Nežárkou (obr. 1). Z důvodů jiných statických objektů a nastavení kamer je nutné při spouštění celé k-means detekce a klasifikace oblaků specifikovat, o jakou instalaci se jedná. Při spuštění *callkmeans.py* skriptu z příkazové řádky²⁷ je nutné zadat parametr určující, pro kterou instalaci systému WILLIAM se má klasifikace tvořit. Při chybně zvoleném parametru samozřejmě dostáváme zkreslené a špatné výsledky. Rozdílnost lokací se projeví také v metodě klasifikace pro východ a západ slunce, které se určují různě podle zeměpisné šířky a délky viz (6.2).

Takto zjednodušeně uvedený popis algoritmu je možné využít v praxi pro samotnou detekci a klasifikaci oblaků a zároveň pro určení pravděpodobnosti srážek. Samozřejmě velký prostor pro zlepšení skýtají také jednotlivé metody klasifikace oblaků, jelikož se zde jedná o prvotní návrh celkového algoritmu. Tím se zabývá kapitola (6.3) společně s již zmiňovaným vyhodnocením výstupů.

6.2 Kritéria hodnocení pravděpodobnosti

Zde budou popsány jednotlivé metody klasifikace, které jsme pro náš vytvořený algoritmus využívali a které generují výstupní pravděpodobnosti skupin oblaků, ke kterým byly dané segmenty snímku přiřazovány.

6.2.1 Pravděpodobnost dle závislosti složky b^* a času od východu / západu slunce

Z výsledků vytvořené analýzy v kapitole 5 jsme přistoupili k nejhodnější získané reprezentaci dat pro nově tvořenou klasifikaci a tou byla závislost chromatické složky b^* na času od východu / západu slunce Δ_s se získanými predikčními intervaly (obr. 18). K výpočtu tedy bylo nutné získat data východů a západů slunce pro lokace, kde se WILLIAM nachází, a s těmi následně pracovat. Pro zvolený parametr lokace při volání klasifikačního algoritmu byl vždy použit přidáný textový soubor obsahující pro danou lokaci uváděné časy. Díky tomu jsme mohli pro vstupní fotografie podle aktuálního času získat časovou vzdálenost od východu či západu slunce. Zároveň k naprogramované klasifikaci byly přidány získané odhady koeficientů pro regresní křivku každé skupiny oblaků společně se šířkou jejich predikčních intervalů. Tyto odhady byly získány z předešlé analýzy v prostředí MATLAB.

²⁷Pro nainstalovaný Python 3 se všemi knihovnamy je možné tyto skripty spustit z terminálu všech operačních systémů.

Vypočtené střední hodnoty segmentů příchozího snímku tak byly následně podle svých barevných hodnot b^* a času vytvoření přiřazeny do této závislosti a porovnávány podle euklidovské vzdálenosti (3.4.0.1). Získání pravděpodobnosti může být provedeno pomocí vzorců následujících dále. Nejprve musela být pro segment $1, \dots, k$ v našem případě pro k rovno 3 (3 výsledné segmenty vstupního snímku), vypočtena euklidovská vzdálenost D_i od regresních křivek jednotlivých skupin oblaků i pro $i = 1, \dots, 4$, která byla dále normována polovinou šířky odhadu predikčního pásu dané skupiny σ_i .

$$\Delta_i^k = \frac{D_i^k}{\sigma_i}. \quad (6.2.1.1)$$

Poté jsme využili inverzní hodnoty vzdálenosti $\delta_i^k = \frac{1}{\Delta_i^k}$ a znormovali ji pro každý segment součtem těchto inverzních vzdáleností ke všem skupinám oblaků.

$$P_i^k = \frac{\delta_i^k}{\sum_i \delta_i^k} \quad (6.2.1.2)$$

Takto jsme pro každý z clusterů (segmentů) získali 4 pravděpodobnosti P , reprezentované procenty, dohromady tedy 12 hodnot. Tím jsme dostali hlavní výstup první metody využívané pro klasifikaci oblaků. V blokovém schématu na obr. 20 je tato metoda zastoupena ve skriptu *classification.py* a na obr. 21 v (zelené) sekci klasifikace společně s dalšími metodami, které budou popsány dále.

6.2.2 Pravděpodobnost dle závislosti složky b^* na elevaci slunce

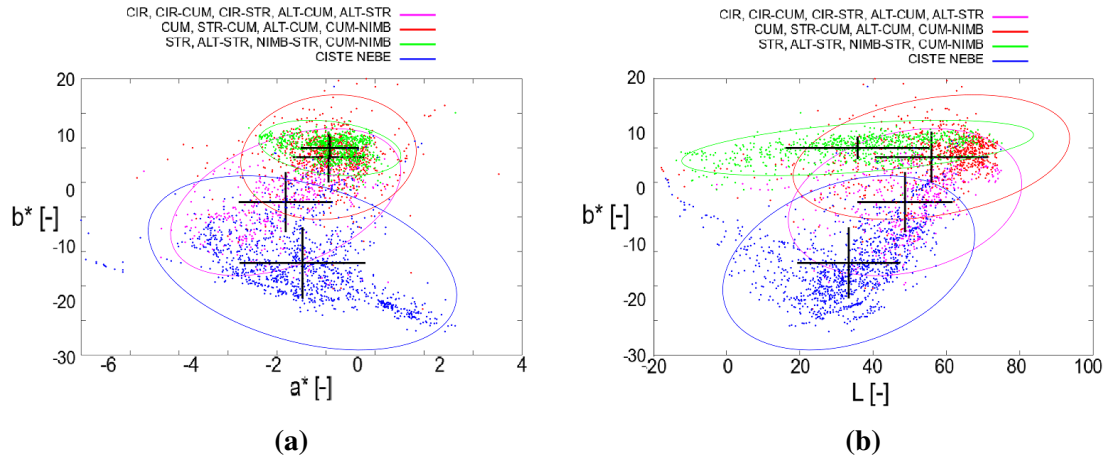
Přestože v předchozí analýze (kapitola 5) nám vyšla závislost b^* na době od východu a západu slunce jako nejvhodnější, rozhodli jsme se využít i závislosti této barevné složky na elevaci slunce. K tomuto rozhodnutí jsme se uchýlili pro získání více pravděpodobnostních hodnot a zpřesnění výsledků, přestože jsme očekávali podobné výstupy jako u předchozí metody. Jedinou změnu ve výpočtu pravděpodobností v tomto případě bylo využití automatického generování elevace slunce pro danou lokaci v určený čas. Zeměpisná délka a šířka instalací systému WILLIAM byla dopředu známá a měli jsme je připravené jako konstanty. Čas byl pokaždé získán z názvu snímku, který tento údaj o době vytvoření obsahoval. Elevace slunce byla počítána automaticky ze zadaných hodnot pomocí Python knihovny *Pysolar*²⁸. Výpočet pravděpodobností poté probíhal stejně jako v případě (6.2.1), pouze v jiném prostoru závislostí.

6.2.3 Pravděpodobnost dle vzdálenosti od chybových elips v prostoru a^* , b^* , respektive L , b^*

Z předchozích grafů 8 a 5b v kapitole 5 jsme zjistili, že jednotlivé podobnostní skupiny oblaků se shlukují do určitých prostorů v těchto dimenzích. Bylo tedy možné získat střední hodnoty těchto prostorů společně se směrodatnými odchylkami a porovnávat v těchto dimenzích vzdálenosti nově příchozích dat od těchto statistik. Jelikož však data v ukázaných grafech byla různorodě natočena, nejvýhodnější bylo středním hodnotám zmíněných skupin (viz 3.1.1) spočítat chybovou elipsu. K tomuto výpočtu byl v prostředí MATLAB využit volně dostupný skript²⁹, který byl podle potřeby zautomatizován a upraven na případ našich dat. Znázornění chybových elips společně se středními hodnotami a odchylkami podobnostních skupin oblaků jsme pro názornost vykreslili na obrázku 6.2.3.

²⁸<https://pysolar.readthedocs.io/>

²⁹<http://www.visiondummys.com/2014/04/draw-error-ellipse-representing-covariance-matrix/>



Obrázek 22: Názorné ukázky chybových elips hodnot z databáze oblaků pro jednotlivé skupiny oblaků společně s černě znázorněnými středními hodnotami a odchylkami dat. Chybové elipsy jsou znázorněny: (a) pro barevné komponenty a^* , b^* , (b) pro komponenty L a b^* .

Hodnoty výsledných elips byly opět uloženy do externích textových souborů a načítány se skripty pro klasifikaci v jazyku Python.

V momentě, kdy jsme měli získané středy shluků skupin ve zvolených dimenzích a hodnoty chybových elips pro tyto skupiny, jsme mohli přistoupit k lokalizaci a tím získání pravděpodobnosti nově přichozících středních hodnot segmentů ze vstupního snímku. Postup výpočtu pravděpodobností vycházel téměř ze stejných rovnic jako v předchozích případech. Jediná změna nastala ve zvoleném Δ_i^k a σ_i ze vzorce (6.2.1.1). Zde uvedené σ již nereprezentovalo predikční interval, ale šířku elipsy od středu prostorového shluku jedné ze skupin oblaků k bodu na elipse, který byl roven minimální vzdálenosti střední hodnoty získaného segmentu v daném dimenzionálním prostoru. Vzdálenost Δ_i^k následně udávala, jak je střední hodnota daného segmentu snímku vzdálena od střední hodnoty prostorového shluku každé ze skupin.

Získané pravděpodobnosti byly opět reprezentovány 12 procentuálními hodnotami, vždy 4 za každou skupinu oblaků přiřazené k jednomu ze 3 segmentů vstupního snímku vytvořených algoritmem k-means (3.4). Jak bylo uvedeno, celý proces této metody probíhal v prostorech L , b^* a a^* , b^* .

Tímto jsme získali pravděpodobnosti výskytu jednotlivých skupin mraků v segmentovaných oblastech na obrázcích ze systému WILLIAM. Tyto procentuální hodnoty každé ze 4 metod klasifikace byly ve výsledku zprůměrovány a nanormovány, aby činily v součtu 100% pro každý segment. Měli jsme proto jednoduchý klasifikační postup, jak detekovat a zároveň určovat s jakou pravděpodobností se dané druhy, respektive skupiny oblaků na snímcích vyskytují. Můžeme konstatovat, že uvedené postupy a metody klasifikace skýtají značný potenciál pro zlepšení, především by zde mohly být přidány další moderní nástroje strojového učení [36, 37], například metody vícenásobné segmentace, SVM (Support-Vector Machine), či neuronové sítě [67].

6.3 Automatická klasifikace v praxi

Jednou ze skupin oblaků, ke které byly získané segmenty snímku přiřazovány, byla i velice často zmiňovaná skupina 3 (ve všech grafech znázorněna zeleně) obsahující srážkové mraky. Jelikož jsme měli vytvořený základní algoritmus pro klasifikaci, mohli jsme určovat, s jakou mírou pravděpodobnosti se tato skupina oblaků na fotografii nachází. Z tohoto závěru bychom mohli zároveň získat nástroj pro určování pravděpodobnosti srážek, což je především hlavní cíl denní evaluace oblohy systému WILLIAM. Dokázali bychom tak vysílat výstrahu při příchodích přívalových deštích a včasné hlásit nebezpečí těchto atmosférických jevů. Nalezený postup skýtá zajisté mnohá úskalí a v budoucnu bude muset být podroben náporu testování a určování pravidel, při jakých rozloženích pravděpodobnosti určitých skupin srážky vznikají. Zároveň bude nutné přesně určit, jak spolehlivě tento model srážky detekuje a s kolika falešnými či neohlášenými detekcemi jsme se setkali. V jádru by se však mohlo jednat o velice zajímavý model detekce srážek, který by mohl představovat značné výhody pro lokální meteorologické předpovědi. K tomuto budou také směřovat příští analýzy.

V současnosti již na zmíněných stránkách projektu WILLIAM³⁰ funguje zde navržená prvotní verze detekce a klasifikace oblaků s výstupy ve formě různých druhů grafů zaměřující se na meteorologické vyhodnocení povětrnostních podmínek či určení pravděpodobnosti deště. Tyto grafy jsou generovány přímo z výsledných dat z klasifikace oblaků především v terminálovém programu Gnuplot³¹.

Ještě před samotnými grafy je na uvedených stránkách dostupný aktuální náhled zkomprimovaného RAW snímku ve formátu JPEG, který byl vytvořen kamerovým systémem. Pro představu zde uvedeme snímek z prosince 2018.

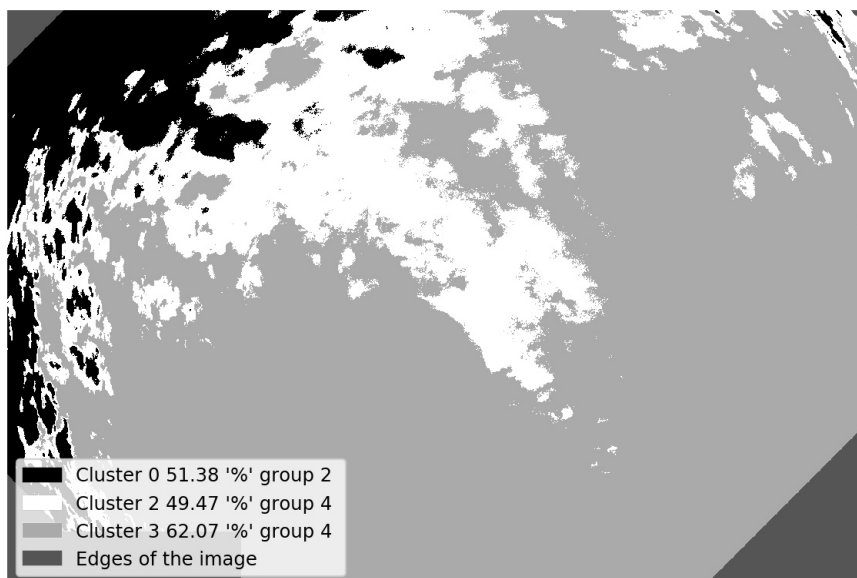


Obrázek 23: Fotografie výstupu ze systému WILLIAM vytvořeného 18. prosince 2018 v 13:20 středoevropského času z náhledu komprimovaného RAW snímku ve formátu JPEG. Na obrázku je možné vidět oblaka nízké a střední oblačnosti společně s čistým nebem [27].

³⁰<http://william.multimediatech.cz/>

³¹<http://www.gnuplot.info/>

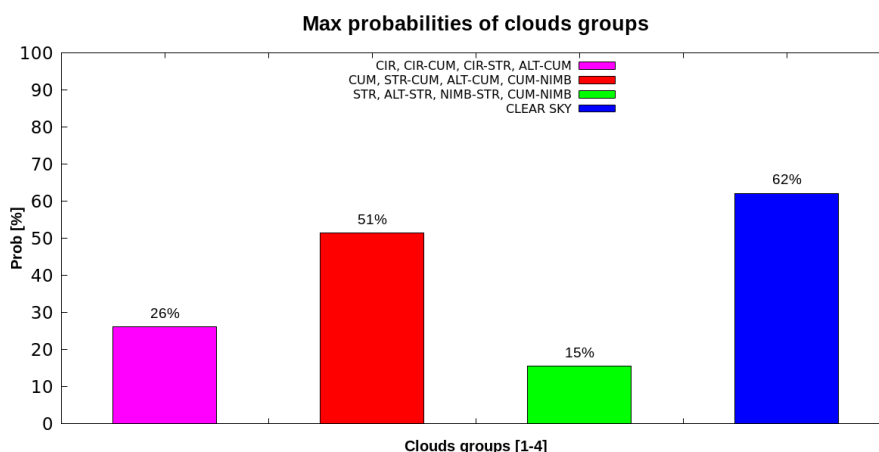
Prvním z generovaných výstupů je ukázka segmentace snímků s přidanou legendou. Ta obsahuje hodnoty určující nejvyšší pravděpodobnost pro určitou skupinu oblaků, číslo skupiny od 1 do 4 (stejně rozložené jako v této práci) a o jaký segment s daným názvem se díky odstínům šedé na snímku jedná (indexovány od 0 do 3). V naznačeném případě jsou mraky nízké oblačnosti na obrázku 6.3 správně detekovány stejně jako segment čistého nebe s vysokými poměrnými pravděpodobnostmi. Častými jevy, se kterými je možné se na denní obloze setkat, jsou značně rozptýlená oblaka neudržující znatelný tvar. V tomto případě cluster 2. Tyto druhy oblačnosti se po barevné transformaci podobají oblačnosti vysoké (na všech grafech fialová) a díky tomu se nacházejí v oblastech komponent prostoru CIELAB mnohem blíže skupině reprezentující čisté nebe. Tento jev může být pozorovatelný i na dále uvedených grafech, a vysvětluje nepřesné přiřazení tohoto segmentu v nastíněném případě.



Obrázek 24: Segmenty výsledného snímku odpovídající jednotlivým skupinám oblaků nalézajících se na vstupním obrázku 23. Pojmenování segmentů vychází z anglického „cluster“ a čísla 0-3 udávající index, se kterým byly vytvářeny. Jeden z clusterů je vždy vyřazen, jelikož se jedná o nechtěné statické okraje snímku. Legenda kromě názvu a barevného přiřazení naznačuje i maximální pravděpodobnost, ke které skupině oblaků (3.1.1) patří.

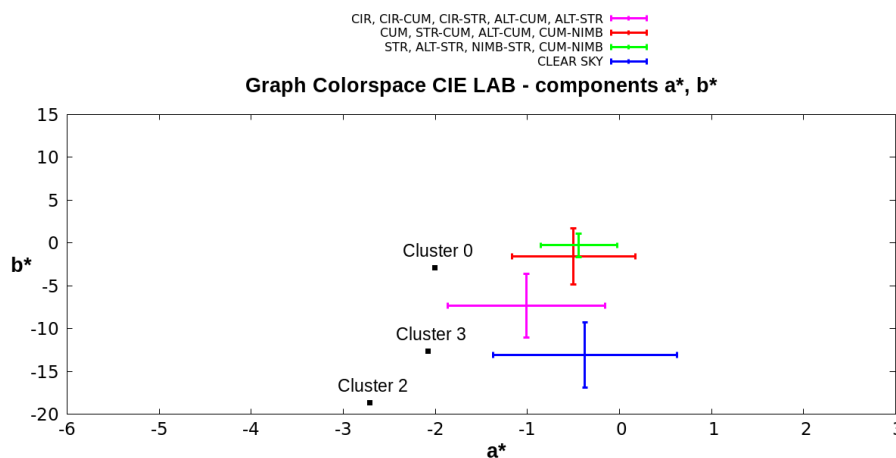
Následujícím z obrázků je graf nejvyšší pravděpodobnosti dané skupiny vyskytující se na aktuálním snímku (obr. 25). Jelikož jsou metody klasifikace postavené na normovaných poměrech vzdáleností tvořených ve výše zmíněných metodách (6.2), žádná ze skupin oblaků nemůže být nulová a tudíž ani stoprocentní, avšak očekáváme, že některé mohou být procentuálně velmi vysoké, pokud splňují klasifikační kritéria.

Na dalších 2 obrázcích 26, 27 jsou znázorněny střední hodnoty segmentů v dimenzích složek barevného prostoru CIELAB. Ačkoliv díky automatické změně měřítka jednotlivých grafů na stránkách systému WILLIAM nemusí být úplně zřejmé, ke kterému předpočítanému středu prostoru skupin oblaků segmenty (clustery) patří, při podrobnějším výzkumu můžeme říct, že nové střední hodnoty segmentů dopadají do předpokládaných oblastí ve zde představených závislostech (8) a (5b). To celkově potvrzuje náš předpo-



Obrázek 25: Maximální pravděpodobnosti přiřazené skupinám oblaků nalézajících se na vstupním rozsegmentovaném snímku (obr. 23).

klad, že je možné detekovat a klasifikovat barevné segmenty našich snímků do určitých skupin, v tomto případě námi navržených. Tento závěr naznačují i výsledné hodnoty jednotlivých pravděpodobností. Máme-li tedy snímek z celooblohové kamery zaměřující se pouze na denní oblohu, je na něm možné detekovat a klasifikovat různé atmosferické jevy, v případě této diplomové práce druhy mraků.

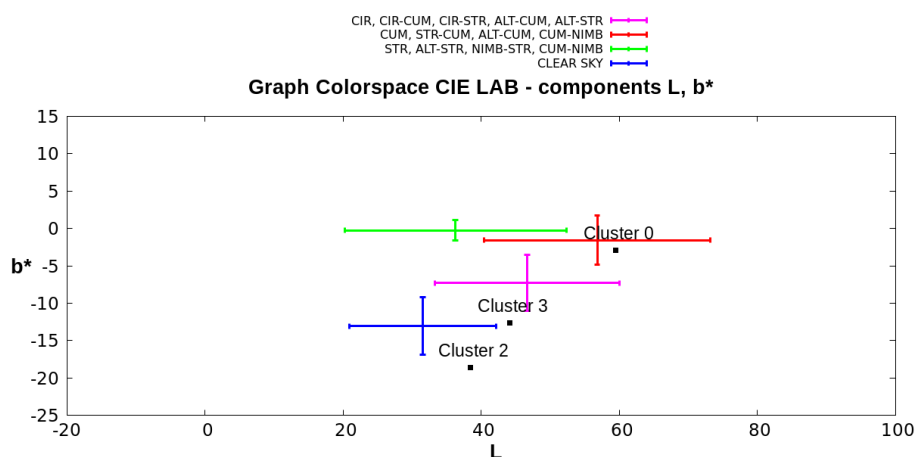


Obrázek 26: Vykreslené střední hodnoty a^* , b^* segmentů (clusterů) vstupního snímku v prostoru uvedených složek. Zároveň jsou zde vyobrazeny střední hodnoty a směrodatné odchylky shluků, kde byly sjednocovány jednotlivé skupiny oblaků z výše představené analýzy.

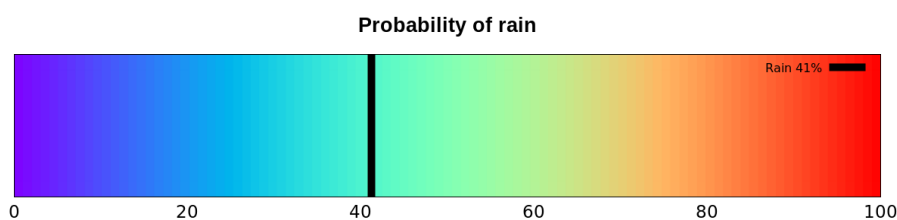
V tomto bodě musíme uvést i potíže, které se při tvorbě celkové klasifikace objevily a které bylo nutné kvůli budoucí funkčnosti vyřešit. Jelikož veškerá předchozí analýza a tvorba databáze probíhala na snímcích z kamery Nikon D5100, respektive softwaru od stejného výrobce Capture NX-D³², byly na snímky (i přes formát RAW) aplikovány algoritmy zlepšení obrazu, o kterých uživatel neměl žádné informace. To se projevilo především při prvotním nasazení zde vytvořené detekce a klasifikace oblaků na novou

³²https://www.nikon.cz/cs_CZ/product/software/capture-nx-d

platformu systému WILLIAM s jiným fotoaparátém. Nově získané body odpovídající středním hodnotám komponent CIELAB jednotlivých segmentů ze vstupního snímku dopadaly v uvedených závislostech L , b^* a a^* , b^* do příliš vzdálených oblastí. Tento problém musel být řešen překalibrací (konverzí a výpočtem nových hodnot) celkové databáze našich snímků pomocí nezávislého programu DCRAW³³, který nad snímky prováděl operace přesně definované uživatelem. Díky tomu bylo možné tyto operace replikovat na veškeré příchozí RAW snímky bez ohledu na typ fotoaparátu či kamery. Tím jsme dosáhli použitelnosti na všechny dostupné verze systému WILLIAM, případně i na jiné širokoúhlé systémy.



Obrázek 27: Vykreslené střední hodnoty L , b^* segmentů (clusterů) vstupního snímku v prostoru uvedených složek. Zároveň jsou zde vyobrazeny střední hodnoty a směrodatné odchylky shluků, kde byly sjednocovány jednotlivé skupiny oblaků z výše představené analýzy.



Obrázek 28: Ukazatel procentuální pravděpodobnosti srážek vytvořený z analýzy aktuálního snímku pomocí algoritmu detekce a klasifikace oblaků. Určení pravidel a testování pro zlepšení přesnosti tohoto indikátoru je v současné době v pokračujícím vývoji.

³³<https://www.cybercom.net/dcoffin/dcrow/>

Posledním a stále vyvíjeným výstupem detekce a klasifikace oblaků je graf znázorněný na obrázku 28. Tento ukazatel by měl být schopný podle složení a detekce jednotlivých skupin oblaků na vstupním snímku indikovat pravděpodobnost příchozích srážek a danou hodnotu znázornit. Protože pro přesné určení pravděpodobnosti dešťových či jiných srážek je potřeba dalších testů, mnohem více dat a nastavení přesných pravidel, kdy je možné hlásit výstrahu, budou všechny tyto kroky v budoucnu analyzovány a zlepšovány. Prvotní reálné výsledky z Jarošova nad Nežárkou však naznačují, že by tento systém mohl fungovat správně a již při prvotním testování zmíněný indikátor podle základních pravidel ukazoval při dešti velmi vysokou pravděpodobnost tvorby srážek. Zlepšení se také může týkat i zmíněného procesu analýzy a předzpracování dat snímků stejně jako použití nových metod strojového učení [36, 37]. Pomocí neuronových sítí by pak v budoucnu mohlo být možné nastavovat pravidla pro určování srážkové pravděpodobnosti automaticky. Následující kroky vývoje představeného systému detekce a klasifikace oblaků, respektive srážek budou zaměřovány a vedeny tímto směrem.

7 Závěr

Celá tato diplomová práce popisuje získávání, postup zpracování a analyzování obrazových dat z celooblohových kamerových systémů, v tomto případě systému WILLIAM. Na této analýze byl následně navržen algoritmus detekce mraků s využitím jejich barevných podobností.

První část této práce (kap. 2) je věnována popisu systému WILLIAM, představení aktuální lokace jeho pozorovacích stanic a nastínění funkčnosti a aplikací, ve kterých je využíván.

Následuje kapitola 3, která se věnuje meteorologické klasifikaci oblaků, v níž jsou popsány barevné prostory CIELAB a CIE XYZ pro detekci těchto atmosferických jevů. Zároveň je zde ukázáno jejich rozdělení do podobnostních skupin. Konečný úsek této kapitoly zaujímá k-means segmentace, která tvořila jádro navrženého algoritmu detekce a klasifikace mraků, a je zde matematicky, společně s jejími modifikacemi, definována.

Jelikož pro tvorbu vhodné klasifikace oblaků bylo nutné zajistit dostatečné množství obrazových dat, je zde popsán postup jejich získávání z celooblohového systému WILLIAM společně s dalším zpracováním. Dále následovalo utvoření rozsáhlé databáze těchto dat (snímků), na které byla aplikována k-means segmentace v různých barevných prostorech a přiřazení takto získaných segmentů k jednotlivým atmosferickým jevům vyskytujícím se na daných snímcích. Následovalo statistické zpracování výpočtů barevných hodnot jednotlivých segmentů a jejich uložení do textových souborů, připravených pro další analýzu. Protože z předchozích rozborů bylo známo, že barevné hodnoty oblaků se v průběhu dne mění, jsou v kapitole 4 představeny vhodné solární parametry, se kterými by bylo možné tyto jevy zkoumat. Jedná se o časovou vzdálenost od pravého poledne, elevaci slunce a čas od východu, respektive západu slunce.

Kapitola 5 se zabývá celkovým rozбором získaných chromatických parametrů atmosferických jevů, které se nalézaly na zpracovaných snímcích. Hned v první provedené analýze byly potvrzeny předpoklady shlukování jednotlivých skupin mraků v chromatických dimenzích prostoru CIELAB. Mezi méně vhodné parametry pro návrh klasifikace oblaků se po podrobnější analýze zařadila barevná složka a^* . Především její vývoj v závislosti na solárních parametrech se pro detekci skupiny mraků jevil jako nevyhovující. Zbylé komponenty L a především pak b^* barevného prostoru CIELAB naopak přijatelné vlastnosti pro návrh algoritmu detekce a klasifikace oblaků prokázaly.

Pro další vyhodnocení závislostí zmiňovaných barevných složek na solárních parametrech byla představena regresní analýza. Ta pomáhala navrhnout analytické průběhy jednotlivých skupin oblaků právě v závislosti na zde představených solárních parametrech. Pro jasovou složku L bylo zjištěno, že všechny skupiny oblaků lze charakterizovat analytickým modelem, kromě skupiny mraků obsahující dešťovou oblačnost, jejíž body vykazovaly rozptyl napříč všemi skupinami. Barevná složka b^* v závislosti na solárních parametrech umožňovala naopak nejpřesnější regresní odhad vývoje hodnot jednotlivých skupin a za pomoci odhadnutých predikčních intervalů bylo možné vymezit úseky, ve kterých nově získaná data náleží určité skupině oblaků s danou pravděpodobností. Toto zjištění nám ze závislosti chromatické komponenty b^* především na elevaci slunce a času od východu a západu slunce umožňovalo vytvořit kritéria pro detekci a klasifikaci oblaků.

Po zevrubném rozboru obrazových dat z vytvořené databáze v předchozích kapitolách se poslední část této diplomové práce (kap. 6) věnuje samotnému návrhu algoritmu pro detekci a klasifikaci mraků. Tento algoritmus byl od základu vyvíjen pro celooblohový systém WILLIAM a vychází z vhodného předzpracování nově získaných snímků, barevné k-means segmentace, statistického zpracování těchto dat, určení pravděpodobnosti výskytu daných atmosferických jevů na vstupních fotografiích podle navržených klasifikačních kritérií a vytvoření adekvátních výstupů. Jako klasifikační metody zde slouží parametry vzdálenosti nově získaných datových bodů od předdefinovaných chybových elips v prostorech L , b^* a a^* , b^* , a metody určující pravděpodobnost podle vypočtených hodnot obrazových segmentů barevné složky b^* v závislosti na elevaci slunce a času od východu, respektive západu slunce.

Naprogramovaný algoritmus byl testován na reálných datech z celooblohové kamery systému WILLIAM, kde byla prokázána jeho funkčnost. V další verzi byl návrh rozšířen také o jednoduchá pravidla, podle kterých bylo možné detekovat a klasifikovat srážkovou oblačnost, což bylo i jedním z našich vytyčených cílů.

Do budoucna, pro značné zpřesnění výsledků detekce typu oblačnosti, bude nutné vylepšit stávající metody klasifikace a vyvinout nové, které bude třeba rozšířit o moderní nástroje strojového učení, například SVM, neuronové sítě či vícenásobnou segmentaci. Taktéž bude zásadní determinovat vliv aberací na přesnost našich výsledků. Obecně však lze konstatovat, že předvedená analýza a detekční algoritmus mraků skýtá značné možnosti v oblasti meteorologie, a bylo by vhodné se v budoucnu zabývat jeho vylepšením.

Seznam použitých zkratk

CIE Commission Internationale d'Edeige.

CMOS Complementary Metal Oxide Semiconductor.

DSLR Digital Single Lens Reflex Camera.

FTP File Transfer Protocol.

GIF Graphics Interchange Format.

GMM Gaussian Mixture Model.

GPS Global Positioning System.

IR Infra Red.

JPEG Joint Photographic Experts Group.

KL Karhunen–Loève transformation.

NEF Nikon Electronic Format.

NOAA National Oceanic and Atmospheric Administration.

PCA Principal Component Analysis.

PSF Point Spread Function.

SSH Secure Shell.

SVM Support-Vector Machine.

TIFF Tagged Image File Format.

WILLIAM Wide-field aLL-sky Image Analyzing Monitoring system.

Seznam obrázků

1	Mapa ČR s lokacemi pro systém WILLIAM	19
2	Schéma pro funkce systému WILLIAM společně s vylepšenou verzí.....	21
3	Ukázka vytvořeného snímku společně s barevnou maskou při předzpracování.....	22
4	Snímek denní oblohy uložený do databáze společně s ukázkou segmentování.	31
5	Předchozí výsledky menší databáze v doménách a^* , b^* s větším obsahem dat z nové databáze	36
6	Předchozí výsledky menší databáze v doménách x , y s větším obsahem dat z nové databáze	37
7	Závislost našich dat v dimenzích L , a^*	38
8	Závislost našich dat v dimenzích L , b^*	38
9	Závislost našich dat v dimenzích L a b^* na elevaci slunce	39
10	Závislost našich dat v dimenzích L a b^* na době od pravého poledne.....	40
11	Závislost našich dat v dimenzích L a b^* na době od východu či západu slunce	41
12	Závislost našich dat pro barevnou komponentu a^* na elevaci slunce, době od pravého poledne a době od východu/západu slunce	42
13	Leave-N-Out cross-validace s naznačenými výsledky pro objektivní metody měření regresních parametrů.	45
14	Regresní křivky pro skupiny oblaků v závislostech L ku elevaci slunce.....	47
15	Regresní křivky pro skupiny oblaků v závislostech L ku času od pravého poledne	48
16	Regresní křivky pro skupiny oblaků v závislostech L ku času od východu /západu slunce	49
17	Regresní křivky pro skupiny oblaků v závislostech b^* ku elevaci slunce	50
18	Regresní křivky pro skupiny oblaků v závislostech b^* ku času od východu / západu slunce	51
19	Střední hodnoty komponenty b^* skupin oblaků pro každou z hodin od pravého poledne	52
20	Diagram detekce a klasifikace oblaků pro skripty v jazyce Python.....	55
21	Diagram pro celý algoritmus detekce a klasifikace oblaků	56
22	Chybové elipsy pro složky L , b^* a a^* , b^*	59
23	Fotografie ze systému WILLIAM, prosinec 2018	60
24	Segmentace vstupní fotografie ze systému WILLIAM s legendou výsledků klasifikace	61
25	Max. pravděpodobnost skupin oblaků vyskytujících se na segmentech snímku	62
26	Výsledky středních hodnot segmentů v prostoru a^* , b^*	62
27	Výsledky středních hodnot segmentů v prostoru L , b^*	63
28	Indikátor pravděpodobnosti srážek	63

Seznam tabulek

1	Maticе koeficientů pro převod z sRGB do XYZ pro zdroj světla D65	24
2	X_n, Y_n, Z_n koeficienty CIELAB prostoru pro zdroj světla D65	25
3	Ukázka výsledné databáze pro CIELAB uložené v programu Excel	33
4	Exportovaná data z vytvořené databáze	34
5	Barevné hodnoty získaných segmentů z vytvořené databáze dat pro CIELAB	35
6	Výsledky barevných hodnot získaných segmentů z vytvořené databáze dat pro CIE XYZ	35
7	Rozšířená ukázka barevných hodnot v prostoru CIELAB z vybraných dnů .	79
8	Rozšířená ukázka barevných hodnot v prostoru CIE XYZ z vybraných dnů	81
9	Delší ukázka z vytvořené databáze oblaků pro CIELAB	83

Literatura

- [1] L. Krauz, “Barevné reprezentace atmosferických jevů ze systému WILLIAM,” bakalářská práce, ČVUT, Fakulta Elektrotechnická, ČVUT v Praze, 2016.
- [2] P. Janout and P. Páta, “Celooblohová kamera s extrémně širokoúhlým zorným polem,” *Jemná mechanika a optika*, vol. 59, no. 10, pp. 281–283, 2014.
- [3] J. Shields, M. Karr, A. Burden, R. W. Johnson, and W. S. Hodgkiss, “Whole sky imaging of clouds in the visible and ir for starfire optical range,” p. 67, 07 2007.
- [4] S. Liu, L. Zhang, Z. Zhang, C. Wang, and B. Xiao, “Automatic cloud detection for all-sky images using superpixel segmentation,” *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 2, pp. 354–358, 2015. [Online]. Dostupne: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6874559>
- [5] J. Adam, J. Buss, K. Brügge, M. Nöthe, W. Rhode, P. Trávníček, M. Prouza, M. Gaug, and B. Keilhauer, “Cloud detection and prediction with all sky cameras,” *EPJ Web of Conferences*, vol. 144, 2017. [Online]. Dostupne: <http://www.epj-conferences.org/10.1051/epjconf/201714401004>
- [6] A. Kazantzidis, P. Tzoumanikas, A. Bais, S. Fotopoulos, and G. Economou, “Cloud detection and classification with the use of whole-sky ground-based images,” *Atmospheric Research*, vol. 113, pp. 80–88, 2012. [Online]. Dostupne: <http://linkinghub.elsevier.com/retrieve/pii/S0169809512001342>
- [7] S. Dev, F. M. Savoy, Y. H. Lee, and S. Winkler, “High-dynamic-range imaging for cloud segmentation,” *Atmospheric Measurement Techniques*, vol. 11, no. 4, pp. 2041–2049, 2018. [Online]. Dostupne: <https://www.atmos-meas-tech.net/11/2041/2018/>
- [8] A. Heinle, A. Macke, and A. Srivastav, “Automatic cloud classification of whole sky images,” *Atmospheric Measurement Techniques*, vol. 3, no. 3, pp. 557–567, 2010. [Online]. Dostupne: <http://www.atmos-meas-tech.net/3/557/2010/>
- [9] W. Zhuo, Z. Cao, and Y. Xiao, “Cloud classification of ground-based images using texture–structure features,” *Journal of Atmospheric and Oceanic Technology*, vol. 31, no. 1, pp. 79–92, 2014. [Online]. Dostupne: <http://journals.ametsoc.org/doi/abs/10.1175/JTECH-D-13-00048.1>
- [10] S. Dev, Y. H. Lee, and S. Winkler, “Color-based segmentation of sky/cloud images from ground-based cameras,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 1, pp. 231–242, 2017. [Online]. Dostupne: <http://ieeexplore.ieee.org/document/7471439/>

- [11] J. Hernández-Andrés, R. L. Lee, and J. Romero, “Calculating correlated color temperatures across the entire gamut of daylight and skylight chromaticities,” *Applied Optics*, vol. 38, no. 27, pp. 5703–, 1999. [Online]. Dostupne: <https://www.osapublishing.org/abstract.cfm?URI=ao-38-27-5703>
- [12] J. Hernández-Andrés, J. Romero, J. L. Nieves, and R. L. Lee, “Color and spectral analysis of daylight in southern europe,” *Journal of the Optical Society of America A*, vol. 18, no. 6, pp. 1325–, 2001. [Online]. Dostupne: <https://www.osapublishing.org/abstract.cfm?URI=josaa-18-6-1325>
- [13] M. Blažek and P. Páta, “Colour transformations and k-means segmentation for automatic cloud detection,” *Meteorologische Zeitschrift*, vol. 24, no. 5, pp. 503–509, 2015-07-16. [Online]. Dostupne: [http://www.schweizerbart.de/papers/metz/detail/prepub/85014/Colour`transformations`and`K`means`segmentation`fo?af=crossref](http://www.schweizerbart.de/papers/metz/detail/prepub/85014/Colour%20transformations%20and%20K%20means%20segmentation%20fo?af=crossref)
- [14] P. Janout, “Modelování prostorově variantních optických systémů,” doktorská práce, ČVUT, Fakulta Elektrotechnická, ČVUT v Praze, 2018.
- [15] P. Páta, P. Janout, M. Blažek, P. Páta, and K. Fliegel, “New generation of meteorology cameras,” in *Photonics, Devices, and Systems VII*. SPIE, 2017-9-27, pp. 55–. [Online]. Dostupne: <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10603/2296846/New-generation-of-meteorology-cameras/10.1117/12.2296846.full>
- [16] P. Tománek, D. Senderáková, P. Páta, P. Janout, P. Páta, J. Bednář, E. Anisimova, M. Blažek, and P. Skala, “Stellar objects identification using wide-field camera,” pp. 94 501I–. [Online]. Dostupne: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2074255>
- [17] A. G. Tescher, E. Anisimova, J. Bednář, M. Blažek, P. Janout, K. Fliegel, P. Páta, S. Vitek, and J. Švihlík, “Estimation and measurement of space-variant features of imaging systems and influence of this knowledge on accuracy of astronomical measurement,” pp. 92 171E–. [Online]. Dostupne: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2061736>
- [18] E. Anisimova, K. Fliegel, M. Blažek, P. Janout, J. Bednář, P. Páta, S. Vitek, J. Švihlík, and A. G. Tescher, “Analysis of images obtained from space-variant astronomical imaging systems,” pp. 885 607–. [Online]. Dostupne: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2023904>
- [19] P. Janout, P. Páta, P. Skala, and J. Bednář, “Psf estimation of space-variant ultra-wide field of view imaging systems,” *Applied Sciences*, vol. 7, no. 2, 2017. [Online]. Dostupne: <http://www.mdpi.com/2076-3417/7/2/151>
- [20] P. Janout, P. Páta, M. B. Levine-West, and M. A. Kahan, “Analysis and performance of non-circular polynomials in the wavefront modelling,” in *Optical Modeling and Performance Predictions X*. SPIE, 2018-9-20, pp. 26–. [Online]. Dostupne: <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10743/2321302/Analysis-and-performance-of-non-circular-polynomials-in-the-wavefront/10.1117/12.2321302.full>

- [21] A. G. Tescher, P. Janout, P. Páta, P. Skala, K. Fliegel, S. Vítek, and J. Bednář, “Application of field dependent polynomial model,” pp. 99710F–. [Online]. Dostupne: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2237310>
- [22] J. Bednar, P. Skala, and P. Pata, “Toward astrometric calibration of ultrawide-field images,” *Astronomische Nachrichten*, vol. 339, no. 5, pp. 403–407, 2018. [Online]. Dostupne: <http://doi.wiley.com/10.1002/asna.201813514>
- [23] A. G. Tescher, K. Fliegel, P. Janout, J. Bednář, L. Krasula, S. Vítek, J. Švihlík, and P. Páta, “Performance evaluation of image deconvolution techniques in space-variant astronomical imaging systems with nonlinearities,” pp. 959927–. [Online]. Dostupne: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?doi=10.1117/12.2187888>
- [24] Nikon, *DIGITAL CAMERA D5100 User’s Manual*, Thailand, 2011. [Online]. Dostupne: <http://cdn-10.nikon-cdn.com/pdf/manuals/dslr/D5100`EN.pdf>
- [25] B. E. Bayer, “Color imaging array,” Patent US3971065A, 1976.
- [26] J. J. Kumler, M. L. Bauer, R. E. Fischer, R. B. Johnson, W. J. Smith, and W. H. Swantner, “Fish-eye lens designs and their relative performance/title,” pp. 360–369. [Online]. Dostupne: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=915923>
- [27] H. Häckel, *Atlas oblaků*. Praha: Academia, 2008.
- [28] P. Dvořák, *Ilustrovaný atlas oblaků*. Cheb: Svět křídel, 2001.
- [29] M. S. Tooms, *Colour reproduction in electronic imaging systems*. Chichester, West Sussex: John Wiley & Sons Inc., [2016].
- [30] T. Smith and J. Guild, “The c.i.e. colorimetric standards and their use,” *Transactions of the Optical Society*, vol. 33, no. 3, pp. 73–134, 1932-01-01. [Online]. Dostupne: <http://stacks.iop.org/1475-4878/33/i=3/a=301?key=crossref.3dec82dbb24569e539ed1a7e2fdbec98>
- [31] D. A. Palmer and A. W. Tarrant, “Cie standard observers,” pp. 6–11. [Online]. Dostupne: <http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=1227411>
- [32] V. Vít, *Televizní technika přenosové barevné soustavy*. Praha: BEN - technická literatura, 1997.
- [33] N. Ohta and A. R. Robertson, *Colorimetry*. Hoboken, NJ, USA: J. Wiley, c2005.
- [34] W. Mendenhall, *Introduction to Probability and Statistics*, 17th ed. Boston: PWS-KENT Publishing Company, 1987.
- [35] R. A. Johnson, I. Miller, and J. E. Freund, *Miller & Freund’s probability and statistics for engineers*, 8th ed. Boston: Prentice Hall, c2011.
- [36] S. Rogers and M. Girolami, *A First Course in Machine Learning*, 2nd ed. United Kingdom: Chapman & Hall/CRC, 2016.

- [37] C. M. Bishop, *Pattern recognition and machine learning*. New York: Springer, c2006.
- [38] S. Lloyd, “Least squares quantization in pcm,” *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982. [Online]. Dostupne: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1056489>
- [39] R. Chellappa, A. Veeraraghavan, N. Ramanathan, C.-Y. Yam, M. S. Nixon, A. Elgammal, J. E. Boyd, J. J. Little, N. Lynnerup, P. K. Larsen, and D. Reynolds, “Gaussian mixture models,” in *Encyclopedia of Biometrics*. Boston, MA: Springer US, 2009, pp. 659–663. [Online]. Dostupne: http://www.springerlink.com/index/10.1007/978-0-387-73003-5_196
- [40] V. Faber, “Clustering and the continuous k-means algorithm,” *Los Alamos Science*, no. Vol.22, pp. 138–144, 1994,.
- [41] G. F. Jenks, “The data model concept in statistical mapping,” in *International Yearbook of Cartography* 7, 1967, p. 186–190.
- [42] A. K. Jain and R. C. Dubes, *Algorithms for clustering data*. Englewood Cliffs, N.J.: Prentice Hall, c1988.
- [43] L. Kaufman and P. Rousseeuw, “Clustering by means of medoids,” in *Statistical Data Analysis Based on the L1-Norm and Related Methods*, North-Holland, 1987, p. 405–416.
- [44] J. C. Dunn, “A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters,” *Journal of Cybernetics*, vol. 3, no. 3, pp. 32–57, 1973. [Online]. Dostupne: <http://www.tandfonline.com/doi/abs/10.1080/01969727308546046>
- [45] T. Kanungo, D. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A. Wu, “An efficient k-means clustering algorithm,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 881–892, 2002. [Online]. Dostupne: <http://ieeexplore.ieee.org/document/1017616/>
- [46] D. Arthur and S. Vassilvitskii, “k-means++: The advantages of careful seeding.” in *SODA '07: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms.*, 2007, pp. 1027–1035.
- [47] I. S. Dhillon, Y. Guan, and B. Kulis, “Kernel k-means,” in *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '04*. New York, New York, USA: ACM Press, 2004, pp. 551–. [Online]. Dostupne: <http://portal.acm.org/citation.cfm?doid=1014052.1014118>
- [48] G. Tzortzis and A. Likas, “The global kernel k-means clustering algorithm,” in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. IEEE, 2008, pp. 1977–1984. [Online]. Dostupne: <http://ieeexplore.ieee.org/document/4634069/>
- [49] S. Basu, I. Davidson, and K. L. Wagstaff, *Constrained clustering*. Boca Raton: CRC Press, c2009.

- [50] R. C. Gonzalez and R. E. Woods, *Digital image processing*, 3rd ed. Upper Saddle River, N.J.: Prentice Hall, c2008.
- [51] F. Seares, *Practical Astronomy for Engineers*. E. W. Stephens publishing Company, 1909. [Online]. Dostupne: <https://books.google.cz/books?id=dHfvAAAAMAAJ>
- [52] W. Millar, *The amateur astronomer's introduction to the celestial sphere*. New York: Cambridge University Press, c2006.
- [53] H. M. Woolf, *On the computation of solar elevation angles and the determination of sunrise and sunset times*. Washington, D.C: National Aeronautics and Space Administration, 1968.
- [54] *The Astronomical Almanac for the Year 2015*. Dept. of the Navy; Annual edition (March 21, 2014), 2015.
- [55] S. A. Glantz and B. K. Slinker, *Primer of applied regression and analysis of variance*. New York: McGraw-Hill, Health Professions Division, c1990.
- [56] G. A. F. Seber and C. J. Wild, *Nonlinear regression*. Hoboken, N.J.: Wiley-Interscience, c2003.
- [57] L. Kubáček and E. Tesaříková, *Weakly nonlinear regression models*, 1st ed. Olomouc: Palacký University, 2008.
- [58] G. Shieh, "Improved shrinkage estimation of squared multiple correlation coefficient and squared cross-validity coefficient," *Organizational Research Methods*, vol. 11, no. 2, pp. 387–407, 2008-03-04. [Online]. Dostupne: <http://journals.sagepub.com/doi/10.1177/1094428106292901>
- [59] K. M. Ramachandran and C. P. Tsokos, *Mathematical statistics with applications*. Boston: Academic Press, c2009.
- [60] K. Pearson, "Liii. on lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, 2010-06-08. [Online]. Dostupne: <https://www.tandfonline.com/doi/full/10.1080/14786440109462720>
- [61] K. Karhunen, *Ueber lineare Methoden in der Wahrscheinlichkeitsrechnung*, ser. Annales Academiae scientiarum Fennicae. Series A. 1, Mathematica-physica, 1947. [Online]. Dostupne: <https://books.google.cz/books?id=Dn3gSAAACAAJ>
- [62] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, no. 6, pp. 417–441, 1933. [Online]. Dostupne: <http://content.apa.org/journals/edu/24/6/417>
- [63] I. T. Jolliffe, *Principal component analysis*, 2nd ed. New York: Springer, 2002.
- [64] J. Bečvář, *Lineární algebra*, vyd. 4 ed. Praha: Matfyzpress, 2010.
- [65] P. Olšák, *Úvod do algebry, zejména lineární*, vyd. 1 ed. Praha: FEL ČVUT v Praze, 2007.

- [66] P. Sovka, *Vybrané metody číslicového zpracování signálů*, vyd. 2 ed. Praha: ČVUT, 2003.
- [67] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [68] R. A. FISHER, “The use of multiple measurements in taxonomic problems,” *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, 1936. [Online]. Dostupne: <http://doi.wiley.com/10.1111/j.1469-1809.1936.tb02137.x>
- [69] P. Comon, “Independent component analysis, a new concept?” *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994. [Online]. Dostupne: <http://linkinghub.elsevier.com/retrieve/pii/0165168494900299>

Příloha A

Tato příloha představuje popis všech adresářů a datových souborů, které obsahuje přiložené CD.

MATLAB analysis

Obsahuje veškeré skripty využívané v této diplomové práci v prostředí MATLAB.

- 1_ORIZNUTI.m - skript pro decimaci TIFF fotografií 800x1200 px
- 2_OREZMASKOU.m - skript pro aplikaci barevné masky na vstupní fotografie
- 3_KMEANS.m - skript pro barevnou transformaci a k-means segmentaci
- 4_UlozCLUSTERdoVYSLEDKU.m - skript pro uložení shluků k-means segmentace
- 5_VykresleniClusteru.m - skript pro vykreslení shluků k-means segmentace
- 6_statikafin.m - skript pro načtení exportovaných dat z databáze (v adresáři *Exported_txt_data*), vypočtení výsledků a jejich uložení do .txt souborů (v adresáři *Exported_results*)
- 7_vykreslvysledky.m - skript pro vykreslení výsledných grafů
- 8_GIF.m - skript vykreslující animované obrázky vyvíjející se podle solárních parametrů v L, b^* a a^*, b^* prostorech
- 9_Fitting.m - skript regresní analýzy našich dat s použitím *curve fitting* toolboxu a naprogramovaná cross-validace
- 10_error_ellipse.m - skript vytvoření chybových elips v L, b^* a a^*, b^* prostorech využitě v programové části detekce oblaků
- 11_PCA.m - skript zabývající se analýzou hlavních komponent solárních parametrů
- errorbarxy.m - skript funkce pro vykreslení grafů se směrodatnými odchylkami
- rsquared.m - pomocný skript s funkcí počítající koeficient determinace
- statika.m - skript pro načtení dat z databáze převedené do .txt souboru, vypočtení výsledků a jejich uložení

Exported results

V tomto adresáři jsou uvedeny výsledky barevných hodnot clusterů pro jednotlivé snímky v barevných prostorech CIELAB (*vysledkyLAB2v2018.txt*) a CIE XYZ (*vysledkyXYZ2v2018.txt*). Společně s těmito soubory se v této složce nachází i textový soubor s popisem jednotlivých sloupců představených dat.

Exported_txt_data

V tomto adresáři se nacházejí textové soubory, exportované z databáze zpracovaných snímků, rozdělené po anotovaných dnech. Celkově se zde nachází 62 .txt souborů.

Gnuplot_graphs

Složka obsahující grafy použité v této diplomové práci.

- runLAB.plt - Gnuplot skript pro vytvoření našich grafů ze vstupních dat v barevném prostoru LAB
- runXYZ.plt - Gnuplot skript pro vytvoření našich grafů ze vstupních dat v barevném prostoru XYZ
- vysledkyLAB2v2018.txt - vstupní data pro Gnuplot skripty v barevném prostoru LAB
- vysledkyXYZ2v2018.txt - vstupní data pro Gnuplot skripty v barevném prostoru XYZ

Images

Obsahuje ukázky zpracovaných snímků uložených do podsložek.

- RAW_NEF - původní surová data ze systému WILLIAM
- TIFF - Exportovaná surová data, převedená na TIFF formát obrazových dat
- TIFF_with_mask - fotografie po aplikování barevné masky odstraňující statické objekty
- CLUSTERS - vysegmentované clustery k-means metodou pro LAB a XYZ rozdělené do stejnojmenných podsložek
- Clouds - příklady typických mraků, které se vyskytují na denní obloze.

Python_algorithm

Zde se nachází výsledná implementace navrženého algoritmu detekce a klasifikace mraků v programovacím jazyce Python.

- 2018_10_30_12_40_08.jpg - testovací snímek pro navržený algoritmus
- callkmeans2.py - hlavní skript zajišťující spuštění veškerých metod
- classification.py - skript obsahující klasifikační metody
- Clusters-means.csv - textový soubor obsahující střední hodnoty shluků skupin mraků v prostorech L , b^* a a^* , b^* pro vykreslení výstupu v programu pro generování grafů Gnuplot s názvem skriptu *plot_LAB_graphs.plt*
- error_ellipseAB.txt - doplňkový .txt soubor obsahující hodnoty chybových elips vypočtených v prostředí MATLAB v prostoru a^* , b^*

- error_ellipseLB.txt - doplňkový .txt soubor obsahující hodnoty chybových elips vypočtených v prostředí MATLAB v prostoru L, b^*
- kmeans2.py - skript obsahující k-means algoritmus a předzpracování obrazových dat
- output_manager.py - skript zajišťující správné výstupy ve formě .txt souborů a grafů
- plot_LAB_graphs.plt - Gnuplot skript pracující s výstupními daty z navrženého algoritmu pro vykreslení vhodných výstupních grafů
- Readme_Version_2.2.0.txt - poznámky ke spuštění Python algoritmu
- Sunset_Jarsov.txt - doplňkový .txt soubor obsahující hodnoty východů a západů slunce pro lokaci systému WILLIAM v Jarošově nad Nežárkou
- Sunset_Prague.txt - doplňkový .txt soubor obsahující hodnoty východů a západů slunce pro lokaci systému WILLIAM v Praze

Database

Zde se nachází výsledné důležité soubory.

- KLASIFIKACE-OBLAK-DATABAZE-V2.xlsx - úplná databáze fotografií pro LAB a XYZ prostory, společně s listy solárních parametrů
- MIDDAY_SUNSET_ELEVATION.xlsx - data solárních parametrů seřazená podle zpracovaných fotografií z vytvořené databáze snímků

Diplomová práce s názvem DP-2019-FEL-KrauzL je uvedena v hlavním adresáři ve formátu .pdf.

Příloha B

Z důvodu velkého množství dat, které se do této diplomové práce nevměstná, byla přidána zde do obsáhlejší vzniklé databáze oblaků, která nicméně představuje stále jen malou část z kompletní databáze. Pro detailní představu o našich analyzovaných datech se v této části nachází rozšířená ukázka výše uvedených tabulek 5 6 vyexportovaných zdrojových dat komponent LAB a XZY, na kterých byla prováděna veškerá analýza. Neuvedené části těchto tabulek a další data jsou doplněny v předchozí příloze A na přílohovém CD.

Tabulka 7: Rozšířená ukázka barevných hodnot v prostoru CIELAB z vybraných dnů, které byly zpracovány v databázi mraků. Tabulka znázorňuje, o jaký den z databáze se jedná, pro jaké číslo snímku daného dne jsou přiřazeny následující hodnoty a skupiny oblaků (3.1.1), kam daný snímek, respektive jeho segmentovaná část patří. Dále se v tabulce nacházejí střední hodnoty bezrozměrných barevných složek \bar{L} , \bar{a}^* , \bar{b}^* , dohromady s jejich standardními odchylkami σ_L , σ_{a^*} , σ_{b^*} . Poslední tři pozice zaujímají již uváděné parametry: časový úsek od pravého poledne Δ_{sn} v hodinách (4.3.1), časová vzdálenost od východu či západu slunce Δ_s v hodinách (4.3.3) a elevace slunce ϕ_e ve stupních (4.3.2).

den	č. fot.	skupina	\bar{L}	σ_L	\bar{a}^*	σ_{a^*}	\bar{b}^*	σ_{b^*}	Δ_{sn}	Δ_s	ϕ_e
[-]	[-]	[-]	[-]	[-]	[-]	[-]	[-]	[-]	[h]	[°]	[h]
...
50	12	2	59.5	12.6	0.0	1.0	-1.6	1.8	0.7	7.4	57.5
50	12	4	43.8	5.3	-0.1	1.5	-13.3	3.5	0.7	7.4	57.5
50	13	3	43.3	11.4	-0.6	0.7	-1.1	1.1	0.9	7.6	56.8
50	14	3	43.2	10.7	-0.7	0.6	-1.0	0.7	1.1	7.6	56.0
50	15	2	52.9	12.0	-0.1	1.0	-2.7	2.7	1.2	7.4	55.0
50	15	4	34.6	4.1	1.6	1.1	-18.5	2.7	1.2	7.4	55.0
50	16	2	68.6	23.1	0.7	1.6	-5.1	5.9	1.4	7.2	54.0
50	16	4	29.3	6.3	2.1	0.9	-20.6	1.1	1.4	7.2	54.0
50	17	2	62.2	20.7	0.5	2.0	-9.4	7.0	1.6	7.1	52.9
50	17	4	36.5	7.6	1.2	0.9	-19.6	1.3	1.6	7.1	52.9
50	18	2	69.1	22.1	0.4	1.6	-8.4	7.2	1.7	6.9	51.8
50	18	4	35.3	7.2	1.2	0.9	-19.4	1.4	1.7	6.9	51.8
50	19	2	65.6	19.4	0.6	1.7	-8.3	6.4	1.9	6.7	49.3
50	19	4	34.5	7.1	1.2	1.0	-19.1	1.6	1.9	6.7	49.3
50	20	2	64.3	17.8	0.3	1.8	-7.2	7.1	2.2	6.4	24.5
50	20	4	35.3	7.6	1.2	0.9	-19.4	1.4	2.2	6.4	24.5
51	1	4	21.4	7.4	1.1	1.2	-16.2	1.6	4.7	1.7	25.8
51	2	4	22.0	7.7	1.2	1.2	-16.6	1.6	4.6	1.8	27.4
51	3	4	22.7	7.9	1.4	1.2	-17.0	1.5	4.4	2.0	29.1
51	4	4	23.3	8.1	1.4	1.2	-17.4	1.5	4.3	2.2	30.7
51	5	4	23.8	8.3	1.5	1.2	-17.7	1.5	4.1	2.3	32.3
51	6	4	24.4	8.6	1.6	1.2	-17.9	1.5	3.9	2.5	33.8
51	7	4	24.7	8.6	1.6	1.1	-18.0	1.4	3.8	2.7	35.4
51	8	4	25.5	8.9	1.6	1.1	-18.1	1.4	3.6	2.8	36.9
51	9	4	26.0	9.2	1.6	1.1	-18.2	1.4	3.4	3.0	38.5

den	č. fot.	skupina	\bar{L}	σ_L	\bar{a}^*	σ_{a^*}	\bar{b}^*	σ_{b^*}	Δ_{sn}	Δ_s	ϕ_e
[-]	[-]	[-]	[-]	[-]	[-]	[-]	[-]	[-]	[h]	[°]	[h]
51	10	4	26.6	9.3	1.7	1.1	-18.3	1.4	3.3	3.2	40.0
51	11	4	27.1	9.7	1.6	1.1	-18.3	1.4	3.1	3.3	41.4
51	12	4	27.8	10.2	1.6	1.1	-18.3	1.4	2.9	3.5	42.9
51	13	4	28.4	10.7	1.6	1.1	-18.3	1.4	2.8	3.7	44.2
51	14	1	76.4	16.4	0.6	1.4	-8.0	4.9	2.6	3.8	45.6
51	14	4	28.7	10.7	1.6	1.1	-18.3	1.4	2.6	3.8	45.6
51	15	4	29.3	10.9	1.6	1.1	-18.3	1.4	2.4	4.0	46.9
51	16	1	78.6	16.0	0.6	1.5	-7.6	5.0	2.3	4.2	48.2
51	16	4	29.9	11.0	1.6	1.1	-18.4	1.3	2.3	4.2	48.2
51	17	1	79.3	16.6	0.7	1.5	-7.4	5.3	2.1	4.3	49.3
51	17	4	30.5	11.2	1.5	1.1	-18.3	1.3	2.1	4.3	49.3
51	18	4	31.1	11.2	1.5	1.1	-18.2	1.3	1.9	4.5	50.5
51	19	4	31.6	11.1	1.5	1.1	-18.2	1.4	1.8	4.7	51.5
51	20	1	42.3	8.7	0.7	1.1	-17.5	1.8	1.6	4.8	52.5
51	20	4	24.5	4.1	2.0	0.8	-18.6	1.0	1.6	4.8	52.5
51	21	1	41.8	8.6	0.7	1.1	-17.7	1.7	1.4	5.0	53.4
51	21	4	24.3	3.9	2.1	0.7	-18.6	0.9	1.4	5.0	53.4
51	22	4	31.8	10.8	1.5	1.1	-18.3	1.3	1.3	5.2	54.2
51	23	4	31.8	10.6	1.6	1.0	-18.3	1.3	1.1	5.3	54.9
51	24	4	32.4	10.6	1.5	1.0	-18.3	1.3	0.9	5.5	55.5
51	25	4	32.2	10.3	1.5	1.0	-18.3	1.3	0.8	5.7	56.0
51	26	4	32.3	10.3	1.5	1.0	-18.3	1.3	0.6	5.8	56.4
51	27	4	32.2	10.1	1.5	1.0	-18.2	1.3	0.4	6.0	56.6
51	28	4	32.3	9.9	1.5	1.0	-18.3	1.3	0.3	6.2	56.7
51	29	4	32.3	9.9	1.5	1.0	-18.2	1.3	0.1	6.3	56.7
51	30	4	32.2	9.5	1.5	1.0	-18.2	1.4	0.1	6.5	56.6
51	31	4	32.1	9.4	1.5	0.9	-18.2	1.4	0.2	6.7	56.4
51	32	4	32.2	9.4	1.5	0.9	-18.2	1.4	0.4	6.8	56.0
51	33	4	32.4	9.3	1.4	0.9	-18.1	1.4	0.6	7.0	55.5
51	34	4	32.3	9.1	1.4	0.9	-18.2	1.4	0.7	7.2	54.9
51	35	4	31.9	8.8	1.5	0.9	-18.3	1.4	0.9	7.3	54.2
51	36	4	31.5	8.6	1.5	0.9	-18.2	1.5	1.1	7.3	53.4
51	37	4	31.3	8.5	1.5	0.9	-18.2	1.4	1.2	7.2	52.5
51	38	4	31.4	8.5	1.5	0.9	-18.2	1.5	1.4	7.0	51.6
51	39	4	31.2	8.3	1.4	0.9	-18.1	1.5	1.6	6.8	50.5
51	40	4	31.1	8.0	1.5	0.9	-18.2	1.5	1.7	6.7	49.4
51	41	4	30.5	7.8	1.5	0.9	-18.2	1.6	1.9	6.5	48.2
51	42	4	30.7	7.8	1.5	0.9	-18.2	1.5	2.1	6.3	46.9
51	43	4	30.6	8.0	1.5	0.9	-18.2	1.5	2.2	6.2	45.6
51	44	4	30.9	8.0	1.4	0.9	-18.2	1.5	2.4	6.0	44.3
51	45	4	30.4	7.8	1.4	0.9	-18.2	1.5	2.6	5.8	42.9
51	46	4	29.8	7.7	1.5	0.9	-18.2	1.5	2.7	5.7	41.5
51	47	4	30.0	7.8	1.5	0.9	-18.3	1.4	2.9	5.5	40.0
51	48	4	29.9	7.8	1.5	0.9	-18.3	1.4	3.1	5.3	38.5
...

Tabulka 8: Rozšířená ukázka barevných hodnot v prostoru CIE XYZ z vybraných dnů, které byly zpracovány v databázi mraků. Tabulka znázorňuje, o jaký den z databáze se jedná, pro jaké číslo snímku daného dne jsou přiřazeny následující hodnoty a skupiny oblaků (3.1.1), kam daný snímek, respektive jeho segmentovaná část patří. Dále se v tabulce nacházejí střední hodnoty bezrozměrných barevných složek \bar{L} , \bar{a}^* , \bar{b}^* , dohromady s jejich standardními odchylkami σ_L , σ_{a^*} , σ_{b^*} . Poslední tři pozice zaujímají již uváděné parametry: časový úsek od pravého poledne Δ_{sn} v hodinách (4.3.1), časová vzdálenost od východu či západu slunce Δ_s v hodinách (4.3.3) a elevace slunce ϕ_e ve stupních (4.3.2).

den	č. fot.	skupina	\bar{X}	σ_X	\bar{Y}	σ_Y	\bar{Z}	σ_Z	Δ_{sn}	Δ_s	ϕ_e
[-]	[-]	[-]	[-]	[-]	[-]	[-]	[-]	[-]	[h]	[°]	[h]
...
27	41	4	0.28	0.05	0.30	0.06	0.42	0.06	1.89	6.90	52.25
27	42	1	0.28	0.06	0.30	0.06	0.43	0.09	2.06	6.73	50.94
27	42	4	0.31	0.26	0.32	0.27	0.37	0.27	2.06	6.73	50.94
27	43	1	0.29	0.25	0.31	0.27	0.40	0.28	2.23	6.57	49.58
27	43	4	0.25	0.01	0.27	0.02	0.48	0.02	2.23	6.57	49.58
27	44	1	0.26	0.04	0.28	0.04	0.47	0.06	2.39	6.40	48.18
27	44	4	0.31	0.26	0.32	0.28	0.37	0.27	2.39	6.40	48.18
27	45	1	0.25	0.01	0.27	0.02	0.48	0.03	2.56	6.23	46.74
27	45	4	0.27	0.05	0.28	0.05	0.45	0.06	2.56	6.23	46.74
27	46	1	0.31	0.37	0.32	0.39	0.37	0.42	2.73	6.07	45.26
27	46	4	0.27	0.05	0.28	0.05	0.45	0.06	2.73	6.07	45.26
27	47	1	0.27	0.05	0.28	0.06	0.45	0.08	2.89	5.90	43.75
27	47	4	0.25	0.01	0.26	0.01	0.49	0.02	2.89	5.90	43.75
27	48	1	0.29	0.26	0.30	0.27	0.41	0.28	3.06	5.73	42.22
27	48	4	0.27	0.03	0.29	0.03	0.44	0.06	3.06	5.73	42.22
27	49	1	0.29	0.25	0.30	0.27	0.41	0.27	3.23	5.57	40.67
27	49	4	0.25	0.01	0.26	0.01	0.49	0.02	3.23	5.57	40.67
27	50	1	0.25	0.01	0.26	0.02	0.49	0.03	3.39	5.40	39.09
27	50	4	0.26	0.04	0.28	0.05	0.46	0.06	3.39	5.40	39.09
27	51	1	0.26	0.05	0.28	0.05	0.46	0.08	3.56	5.23	37.51
27	51	4	0.24	0.01	0.26	0.01	0.49	0.02	3.56	5.23	37.51
27	52	1	0.30	0.36	0.32	0.37	0.38	0.40	3.73	5.07	35.90
27	52	4	0.24	0.01	0.26	0.01	0.50	0.02	3.73	5.07	35.90
27	53	1	0.31	0.28	0.32	0.29	0.37	0.29	3.89	4.90	34.29
27	53	4	0.27	0.05	0.28	0.05	0.45	0.06	3.89	4.90	34.29
27	54	1	0.27	0.05	0.29	0.05	0.44	0.09	4.06	4.73	32.67
27	54	4	0.27	0.17	0.29	0.18	0.44	0.19	4.06	4.73	32.67
27	55	1	0.24	0.01	0.26	0.01	0.50	0.02	4.23	4.57	31.04
27	55	4	0.29	0.24	0.30	0.25	0.41	0.26	4.23	4.57	31.04
27	56	1	0.24	0.01	0.26	0.01	0.50	0.02	4.39	4.40	29.41
27	56	4	0.30	0.35	0.32	0.36	0.38	0.39	4.39	4.40	29.41
27	57	1	0.28	0.07	0.29	0.08	0.43	0.12	4.56	4.23	35.97
27	57	4	0.25	0.04	0.27	0.04	0.48	0.06	4.56	4.23	35.97
28	1	1	0.28	0.06	0.30	0.06	0.42	0.07	3.95	2.95	37.58
28	1	4	0.30	0.05	0.32	0.06	0.38	0.07	3.95	2.95	37.58
28	2	1	0.28	0.09	0.30	0.09	0.42	0.10	3.78	3.12	39.18

den	č. fot.	skupina	\bar{X}	σ_X	\bar{Y}	σ_Y	\bar{Z}	σ_Z	Δ_{sn}	Δ_s	ϕ_e
[-]	[-]	[-]	[-]	[-]	[-]	[-]	[-]	[-]	[h]	[°]	[h]
28	2	4	0.28	0.04	0.30	0.04	0.41	0.05	3.78	3.12	39.18
28	3	1	0.31	0.32	0.33	0.33	0.37	0.37	3.62	3.28	40.77
28	3	4	0.27	0.02	0.29	0.02	0.43	0.02	3.62	3.28	40.77
28	4	1	0.31	0.28	0.33	0.29	0.37	0.31	3.45	3.45	42.34
28	4	4	0.28	0.02	0.30	0.02	0.42	0.02	3.45	3.45	42.34
28	5	1	0.30	0.07	0.32	0.08	0.38	0.09	3.28	3.62	43.88
28	5	4	0.28	0.02	0.30	0.02	0.43	0.02	3.28	3.62	43.88
28	6	1	0.31	0.31	0.32	0.32	0.37	0.34	3.12	3.78	45.41
28	6	4	0.31	0.05	0.32	0.06	0.36	0.07	3.12	3.78	45.41
28	7	1	0.29	0.06	0.31	0.07	0.40	0.08	2.95	3.95	46.91
28	7	4	0.31	0.05	0.32	0.06	0.37	0.08	2.95	3.95	46.91
28	8	1	0.29	0.04	0.31	0.05	0.40	0.05	2.78	4.12	48.37
28	8	4	0.31	0.17	0.33	0.17	0.36	0.19	2.78	4.12	48.37
28	9	1	0.28	0.03	0.30	0.03	0.42	0.04	2.62	4.28	49.80
28	9	4	0.30	0.06	0.32	0.06	0.38	0.07	2.62	4.28	49.80
28	10	1	0.29	0.07	0.31	0.07	0.40	0.09	2.45	4.45	51.19
28	10	4	0.30	0.21	0.32	0.21	0.37	0.21	2.45	4.45	51.19
28	11	1	0.30	0.05	0.32	0.05	0.37	0.06	2.28	4.62	52.54
28	11	2	0.30	0.03	0.32	0.04	0.38	0.03	2.28	4.62	52.54
28	11	4	0.31	0.19	0.33	0.19	0.37	0.20	2.28	4.62	52.54
28	12	2	0.31	0.35	0.33	0.37	0.37	0.39	2.12	4.78	53.82
28	12	4	0.30	0.04	0.32	0.04	0.37	0.04	2.12	4.78	53.82
28	13	1	0.31	0.07	0.33	0.08	0.36	0.08	1.95	4.95	62.11
28	13	2	0.28	0.03	0.30	0.03	0.41	0.03	1.95	4.95	62.11
28	13	4	0.32	0.18	0.33	0.19	0.35	0.21	1.95	4.95	62.11
28	14	1	0.31	0.25	0.33	0.26	0.36	0.28	0.22	7.12	61.85
28	14	2	0.32	0.10	0.33	0.11	0.35	0.12	0.22	7.12	61.85
...

Tabulka 9: Tabulka na následujících stranách představuje větší část vytvořené databáze, z tabulkového editoru Excel, všech 2044 snímků. První 2 sloupce značí číslo z celkového počtu fotografií a ID - název jednotlivých obrazových souborů. Dalšími parametry jsou číslo pořizovacího snímku pro daný den, datum a čas. Sloupec *Cluster se sluncem* značí barevný segment zpracovávaného snímku, na kterém je zobrazeno slunce. Následuje 11 sloupců (druhů mraků 3.1.1), kde je subjektivně znázorněno, k jakému typu jednotlivé clustery z k-means barevné segmentace patří. Sloupec *Edges* označuje, který ze segmentů jsou nechtěné okraje snímku. Rain binárně představuje stav kamery, buď při srážkách rovný 1 či bez srážek rovný 0. Nejdůležitějšími částmi této databáze jsou poslední parametry: přiřazení clusterů ke skupinám oblaků 3.1.1, čas od pravého poledne Δ_{sn} v hodinách, respektive od západu a východu slunce Δ_s taktéž v hodinách a elevace slunce ϕ_e ve stupních.

Tabulka 9

Č. řádky	ID	no.	Datum	Čas	Cluster se sluncem	Čivě nebe	Cumulus	Stratus	Stratocumulus	Nimbostratus	Altostratus	Cumulo nimbus	Cirrocumulus	Cirrus	Edges	Rain	1.	2.	3.	4.	Δ_{5m} [h]	Δ_{10} [h]	Φ_e [°]
1809	2016.08.18_15.00_37.nef	1	18.8.2016	15:20	3	1	3							4	2	0	4	3	1	3	3,71444	4,85000	37,2764155
1810	2016.08.18_15.10_07_+1.nef	2	18.8.2016	15:10	4	1	34								2	0		34	1	1	3,10778	5,01667	35,8016032
1811	2016.08.18_15.20_07_+1.nef	3	18.8.2016	15:20	1	4	1					34			1	0	34	1	4	4	3,71444	4,85000	34,3002832
1812	2016.08.18_15.30_08_+1.nef	4	18.8.2016	15:30	1	3	1					34			2	0	34	1	3	3	3,44111	4,68333	32,7739776
1813	2016.08.18_15.40_07_+1.nef	5	18.8.2016	15:40	3	1	3					34			2	0	34	1	3	3	3,60778	4,51667	31,2574736
1814	2016.08.18_15.50_07_+1.nef	6	18.8.2016	15:50	1	3	1					34			2	0	34	1	3	3	3,71444	4,35000	29,6384499
1815	2016.08.18_16.00_07_+1.nef	7	18.8.2016	16:00	1	3	1					4			2	0	4	1	3	3	3,94111	4,18333	28,0797345
1816	2016.08.18_16.10_07_+1.nef	8	18.8.2016	16:10	3	1	3					4			2	0	4	3	1	1	4,10778	4,01667	26,4771174
1817	2016.08.18_16.20_07_+1.nef	9	18.8.2016	16:20	3	1	3					4			2	0	4	3	1	1	4,27444	3,85000	24,8679235
1818	2016.08.18_16.30_06_+1.nef	10	18.8.2016	16:30	1	4	1					3			2	0	3	1	4	4	4,44111	3,68333	23,2493832
1819	2016.08.18_16.40_08_+1.nef	11	18.8.2016	16:40	4	1	4					3			3	0	3	4	1	4	4,60778	3,51667	21,6299933
1820	2016.08.18_16.50_07_+1.nef	12	18.8.2016	16:50	4	12	4								3	0	4	4	12	14	4,77444	3,35000	19,9925674
1821	2016.08.18_17.00_07_+1.nef	13	18.8.2016	17:00	3	14	3								2	0	1	3	14	14	4,94111	3,18333	18,3582175
1822	2016.08.18_17.10_07_+1.nef	14	18.8.2016	17:10	3	14	3								2	0	3	3	14	14	5,10778	3,01667	16,7223974
1823	2016.08.18_17.20_07_+1.nef	15	18.8.2016	17:20	2	14	2								3	0	2	2	14	14	5,27444	2,85000	15,0883934
1824	2016.08.18_17.30_07_+1.nef	16	18.8.2016	17:30	4	3	4								1	2	0	1	3	3	5,44111	2,68333	13,4524847
1825	2016.08.18_17.40_08_+1.nef	17	18.8.2016	17:40	1	4	1								3	2	0	3	1	4	5,60778	2,51667	11,8218188
1826	2016.08.18_17.43_38.nef	18	18.8.2016	17:42	1	4	4								3	2	0	3	4	1	5,64111	2,48333	10,1425572
1827	2016.08.30_07.53_34.nef	1	30.8.2016	7:51	1	34									2	0	1	34	1	34	4,13866	1,66667	27,3379661
1828	2016.08.30_08.10_07_+1.nef	2	30.8.2016	8:10	4	1									2	0	34	1	34	1	3,83866	1,96667	29,6231344
1829	2016.08.30_08.20_07_+1.nef	3	30.8.2016	8:20	1	34									2	0	1	34	1	34	3,67139	2,13333	30,5626889
1830	2016.08.30_08.30_07_+1.nef	4	30.8.2016	8:30	1	4	4								2	0	13	4	4	4	3,9472	2,30000	32,6561018
1831	2016.08.30_08.40_08_+1.nef	5	30.8.2016	8:40	3	14	3								2	0	3	3	14	14	3,33866	2,46667	33,4793253
1832	2016.08.30_08.50_07_+1.nef	6	30.8.2016	8:50	4	13	4								2	0	4	4	13	13	3,17139	2,63333	34,8891672
1833	2016.08.30_09.00_07_+1.nef	7	30.8.2016	9:00	3	24	3								1	0	3	3	24	24	3,0472	2,80000	36,1618191
1834	2016.08.30_09.10_07_+1.nef	8	30.8.2016	9:10	4	23	4								1	0	4	4	23	23	2,83866	2,96667	37,5929937
1835	2016.08.30_09.20_07_+1.nef	9	30.8.2016	9:20	3	14	3								2	0	3	3	14	14	2,67139	3,13333	47,6803148
1836	2016.08.30_10.00_07_+1.nef	10	30.8.2016	11:00	3	14	3								2	0	3	3	14	14	1,0472	4,80000	48,1354568
1837	2016.08.30_11.00_07_+1.nef	11	30.8.2016	11:10	2	14	2								3	0	2	2	14	14	0,83866	4,96667	48,7086135
1838	2016.08.30_11.20_08_+1.nef	12	30.8.2016	11:20	1	24	1								3	0	1	24	1	24	0,67139	5,13333	49,0822318
1839	2016.08.30_11.30_08_+1.nef	13	30.8.2016	11:30	3	14	3								2	0	3	3	14	14	0,50472	5,30000	49,1512945
1840	2016.08.30_11.40_07_+1.nef	14	30.8.2016	11:40	3	14	3								2	0	3	3	14	14	0,33866	5,46667	47,6792447
1841	2016.08.30_13.00_07_+1.nef	15	30.8.2016	13:00	1	34	1								2	0	1	34	1	34	0,99528	6,78333	46,9967362
1842	2016.08.30_13.10_07_+1.nef	16	30.8.2016	13:10	1	3	14								2	0	14	3	14	3	1,15194	6,61667	46,1543182
1843	2016.08.30_13.20_07_+1.nef	17	30.8.2016	13:20	1	34	1								2	0	1	34	1	34	1,32861	6,45000	45,393844
1844	2016.08.30_13.30_08_+1.nef	18	30.8.2016	13:30	1	34	1								2	0	1	34	1	34	1,48528	6,28333	44,4742818
1845	2016.08.30_13.40_07_+1.nef	19	30.8.2016	13:40	4	13	4								2	0	4	4	13	13	1,65194	6,11667	43,4808211
1846	2016.08.30_13.50_07_+1.nef	20	30.8.2016	13:50	1	34	1								2	0	1	34	1	34	1,82861	5,95000	42,4186446
1847	2016.08.30_14.00_07_+1.nef	21	30.8.2016	14:00	1	34	1								2	0	1	34	1	34	1,99528	5,78333	41,3928109
1848	2016.08.30_14.10_07_+1.nef	22	30.8.2016	14:10	3	14	3								2	0	3	3	14	14	1,15194	5,61667	40,1082364
1849	2016.08.30_14.20_07_+1.nef	23	30.8.2016	14:20	3	14	3								2	0	3	3	14	14	1,32861	5,45000	38,8696701