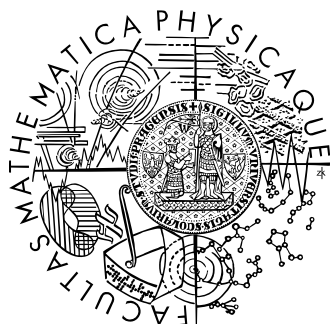


Univerzita Karlova v Praze  
Matematicko-fyzikální fakulta

# DIPLOMOVÁ PRÁCE



Bc. Petr Šácha

## Možnosti využití oblačných produktů družicových dat

Katedra meteorologie a ochrany prostředí

Vedoucí diplomové práce: RNDr. Petr Pišoft, Ph.D.

Studijní program: Fyzika, obor Meteorologie a klimatologie

2012

Mé největší poděkování patří vedoucímu práce RNDr. Petru Pišoftovi, Ph.D. a konzultantovi Mgr. Michalu Žákovi, Ph.D. za vedení mé diplomové práce a hlavně za obrovskou spoustu času, kterou mi věnovali. Bez jejich rad a dobře mířených poznámek by tato práce jen těžko vznikala. Zapomenout nemohu ani na Mgr. Jiřího Mikšovského, Ph.D., který přispěl radami ve statistické oblasti práce. Mé díky patří také rodině, přítelkyni a všem přátelům, neboť veškeré síly potřebné pro tvůrčí aktivitu jsem sbíral právě s nimi. Vám všem tedy hluboce děkuji. Chtěl bych věnovat tuto práci vzpomínce na Vladimíra Šácha staršího.

Prohlašuji, že jsem svou diplomovou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce a jejím zveřejňováním.

V Praze dne 2.8.2012

Bc. Petr Šácha

# Obsah

Úvod .....	1
<b>1 Základní informace .....</b>	<b>3</b>
1.1 Distanční pozorování .....	3
1.1.1 Polární družice.....	4
1.1.2 Geostacionární družice .....	6
1.2 EUMETSAT a síť „SAFů“ .....	8
1.3 CM-SAF.....	9
1.4 Oblačné produkty .....	14
1.4.1 CFC (Cloud Fractional Cover) .....	15
1.4.2 CTY (Cloud Type) .....	18
<b>2 Analyzovaná data a použité statistické metody.....</b>	<b>20</b>
2.1 Soubory dat a koncept validace.....	20
2.1.1 Oblačné produkty družicových dat.....	20
2.1.2 Referenční staniční měření .....	21
2.1.3 Příprava dat .....	22
2.2 Statistické metody .....	26
2.2.1 Spojité veličiny.....	27
2.2.2 Diskrétní veličiny .....	28
<b>3 Analýza denního produktu CFC v oblasti ČR .....</b>	<b>31</b>
3.1 Základní srovnání řady CFC a referenční řady .....	31
3.2 Prostorová variabilita shody řady CFC a referenční řady .....	36
3.3 Časová závislost kvality produktu CFC .....	42
3.4 Závislost kvality vybraných družicových produktů na typu oblačnosti .....	48
<b>Diskuse .....</b>	<b>55</b>
<b>Závěr .....</b>	<b>58</b>
<b>Seznam použitých zkratk.....</b>	<b>61</b>
<b>Literatura a použité internetové odkazy .....</b>	<b>65</b>

**Název práce:** *Možnosti využití oblačných produktů družicových dat*

Autor: *Bc. Petr Šácha*

Katedra: *Katedra meteorologie a ochrany prostředí*

Vedoucí diplomové práce: *RNDr. Petr Pišoft, Ph.D.*

e-mail vedoucího: [Petr.Pisoft@mff.cuni.cz](mailto:Petr.Pisoft@mff.cuni.cz)

**Abstrakt:** Oblačnost hraje zásadní roli v globální energetickém a hydrologickém cyklu. Konkrétně přítomnost oblaků dominuje planetárnímu albedu a účastní se mnoha klimatických zpětných vazeb. V této práci je nejprve podán krátký informativní přehled distančního pozorování, základních informací o organizaci EUMETSAT, jí využívaných družic a její části CM-SAF a rešerše současného výzkumu. Poté je pozornost zaměřena na oblačné produkty získané z družicových dat, speciálně na CFC (procentuelní zakrytí oblačností) a CTY (typ oblačnosti) produkty. Množiny dat denních průměrů těchto produktů jsou porovnávány s denními průměry množství a typu oblačnosti vytvořenými ze staničních SYNOP měření na území České republiky. V případě velkých odchylek mezi zkoumanými datovými soubory denních průměrů pokrytí oblačností jsou hledány možné příčiny jako výběrová chyba, závislost na ročním období, na umístění stanice a na typu a výšce oblačnosti. Je vypočteno mnoho statistických analýz a validačních kritérií. Nakonec jsou diskutovány možnosti využití zkoumaných družicových produktů.

**Klíčová slova:** oblačnost, distanční pozorování, EUMETSAT, CM-SAF, Cloud Fractional Cover, Cloud Type

**Title** *Usage possibilities of the cloud products of satellite datasets*

Author: *Bc. Petr Šácha*

Department: *Department of Meteorology and Environment Protection*

Supervisor: *RNDr. Petr Pišoft, Ph.D.*

Supervisor's e-mail address: [Petr.Pisoft@mff.cuni.cz](mailto:Petr.Pisoft@mff.cuni.cz)

**Abstract:** Cloudiness plays an important role in the global energy and water cycle. In particular, the presence of clouds dominates the planetary albedo and takes part in many climate feedback processes. In this thesis a short informational overview of remote sensing, a description of EUMETSAT, satellites, which it used, its part CM-

SAF and a search retrieval of current research is given at first. Then the study is focused on the cloud satellite products, especially on CFC (cloud fractional cover) and CTY (cloud type) products. Data sets of daily averages of these products are compared with the daily averages created from the surface SYNOP observations of the total cloud cover and cloud type in the area of the Czech Republic. In the case of big variances between the two examined datasets of daily averaged cloud coverage, possible causes like the sampling error, dependence on season, localization of the station and the elevation and type of cloudiness, are searched. Several statistic analyses and validation scores are computed. Finally, possibilities of the examined satellite products are discussed.

**Keywords:** cloudiness, remote sensing, EUMETSAT, CM-SAF, CFC, CTY

# Úvod

Lidstvo bylo odedávna fascinováno přírodou, světem a procesy, které se v jemu známém prostředí odehrávaly. Toto nadšení spolu s touhou poznávat a pochopit se nijak nezmenšovalo s rostoucím stupněm poznání, spíše naopak. Díky stále novým hádankám a podivně se chovajícím systémům vynořujícím se, ať už při zkoumání samé podstaty věcí, či při možnosti komplexnějšího pohledu na ně, musí lidstvo stále opravovat své staré teorie, nebo je dokonce nahrazovat novými modely fungování světa.

V současnosti je jedním z aktuálních a nejvíce diskutovaných vědeckých témat studium zemského klimatu, jeho variability a zhodnocení vlivu člověka. Při studiu klimatu Země, jakožto úplného systému, je ovšem potřebné pozorovat ho jako celek v globálním měřítku, přičemž jediným způsobem, jak toto uskutečnit, jsou distanční pozorování z vesmíru (Reuter et al., 2010).

V této práci se budeme zabývat oblačnými produkty získanými z družicových dat a možnostmi jejich využití v klimatologii. Motivace ke studiu je zřejmá, oblačnost hraje důležitou roli v koloběhu vody a globálním energetickém cyklu. Jmenovitě oblačnost dominuje planetárnímu albedu, které je rozhodující veličinou pro množství slunečního záření odraženého zpět do kosmu. Pro studium vývoje klimatu je tudíž velmi důležité dokumentovat změny v planetární odrazivosti spojené se změnami globální oblačnosti (kvantitativními i kvalitativními) a objevovat potenciální zpětné vazby vzhledem k teplotním změnám či obsahu vodní páry v atmosféře (Karlsson a kol., 2011). Připočteme-li, že různé typy oblaků mají různé charakteristiky odrazivosti a absorpce radiace, potřebujeme ke studiu klimatu nejen znalosti o celkovém množství oblačnosti, ale i o jejím typu, optické tloušťce, situaci na její horní hranici atd. Jak jsou tyto požadavky uspokojovány bude diskutováno dále v práci.

Motivací pro studium oblačnosti je dále její velká spjatost s dopadajícím slunečním zářením na zemský povrch. Dopadající záření je hnací silou všech atmosférických procesů a svou interakcí s rozličnými přenašeči určuje distribuci energie na Zemi (Träger-Chatterjee a kol., 2010). Znalost množství dopadajícího slunečního záření je nezbytná pro mnohá odvětví lidské činnosti jako zemědělství, lesnictví a samozřejmě pro živě diskutované odvětví solární energetiky. Tato odvětví potřebují spíše informace na lokální úrovni, a ač jsou dopadající solární záření i

různé oblačné produkty měřeny i na přízemních pozorovacích stanicích, ukazují se, jak uvádí Journee a kol.(2010), tato pozorování pro mnoho aplikací prostorově i časově nedostatečná. Proto je potřeba jejich konfrontace s družicovými daty.

Předkládaná práce je členěna následovně. V úvodní teoretické části je podáno několik základních informací o distančním pozorování a meteorologických družicích, nezabíhali jsme ovšem do výkladu fyzikálních zákonitostí a základů tohoto odvětví. Případně zájemce je možno odkázat na námi často citovanou knihu Řezáčové (2007) a na monografii Bednáře (1989), které mohou posloužit k uvedení do problematiky šíření elektromagnetických vln v atmosféře, na učebnici cele věnovanou družicové meteorologii R.R. Kelkara (2007) a případně na obsáhlou a teoreticky zaměřenou knihu M. Capderou (2005). Dále bude představena organizace EUMETSAT, která hraje v Evropě vedoucí roli v oblasti distančního pozorování. Nastíníme její cíle a činnost, čímž se dostaneme k síti tzv. SAFů (zkratka ze slov Satellite Application Facility). Ty se zabývají aplikací družicových dat a tvorbou specializovanějších produktů. Největší pozornost bude věnována SAFu zaměřenému na monitorování zemského klimatu, jímž je CM-SAF. Uvedeme výčet jeho produktů a detailně rozebereme námi používané produkty CFC (procentuální pokrytí oblačností) a CTY (typ oblačnosti).

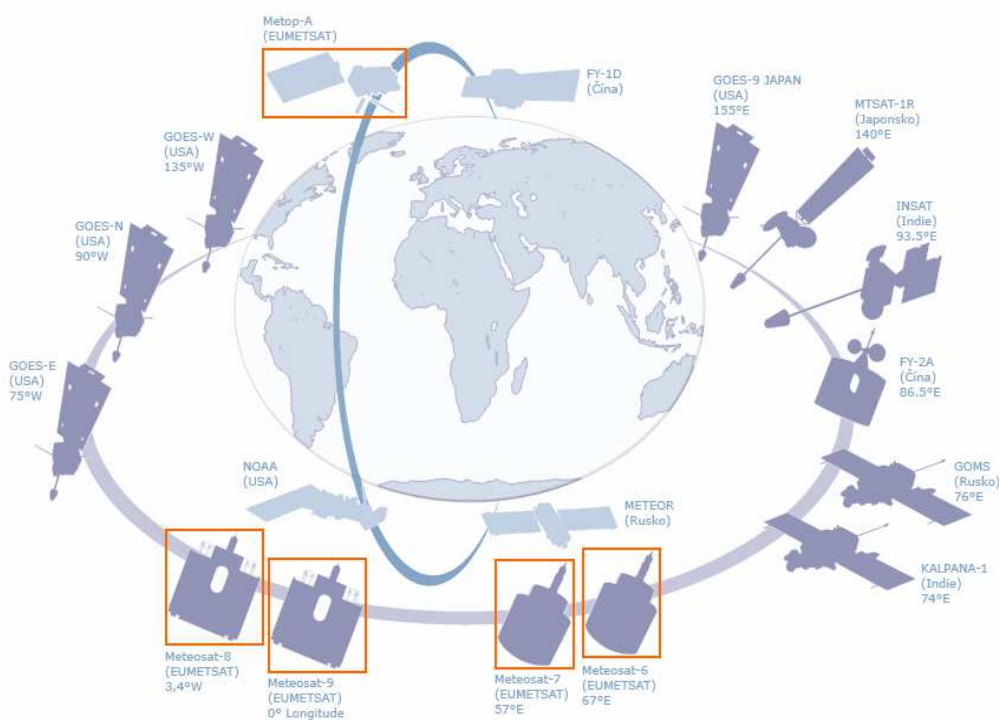
Druhá část práce je věnována analýze denních průměrů výše jmenovaných operativně poskytovaných produktů, přičemž konečným cílem je rozvážit použitelnost oblačných produktů. Dílčími cíli jsou ověření kvality dat v závislosti např. na ročním období, na geografické poloze, na typu oblačnosti atd. Pro porovnání byly k dispozici záznamy přízemních meteorologických pozorování veličin množství a druhu oblačnosti z devíti stanic na území České republiky, které (i přes své známé přesnostní limity) byly brány jako vyjádření reálného stavu atmosféry. Z těchto staničních dat byly vypočítány denní průměry, načež byly sestaveny co nejlépe si prostorově a časově odpovídající řady srovnávaných veličin. Ty byly statisticky analyzovány a jejich shodnost byla testována také množstvím validačních kritérií.

Výsledky jsou prezentovány logicky utříděně v návaznosti na hypotézy o vlastnostech a shodnosti řad. Vzhledem k některým překvapivým výsledkům jsou formulovány hypotézy nové. V závěru jsou pak diskutovány nejistoty prezentovaných výsledků, smysluplnost postupu a je hodnoceno splnění cílů.

# 1 Základní informace

## 1.1 Distanční pozorování

První meteorologická družice neboli specializovaná umělá družice Země zaměřená na monitorování zemské atmosféry byla vypuštěna na nízkou polární dráhu roku 1960 (TIROS-1, USA). Na vypuštění družice na obtížněji dosažitelnou geostacionární dráhu se čekalo až do roku 1966 (ATS-1, USA) a první evropskou družicí se stal Meteosat-1 vypuštěný na geostacionární dráhu roku 1977. Od té doby bylo na oběžnou dráhu vyneseno mnoho meteorologických družic. To je schematicky ilustrováno na obr. 1.1, kde je znázorněna poloha několika významných polárních i geostacionárních družic.



*Obr. 1.1: Globální systém meteorologických družic na přelomu 20. a 21. století (převzato od ČHMÚ)*

Počátečním úkolem meteorologických družic bylo především monitorování atmosféry v globálním měřítku. Ovšem díky zlepšující se geometrické rozlišovací schopnosti, exaktnější kalibraci a schopnosti snímat různá spektrální pásma je možno z družicových pozorování získávat stále pokročilejší a detailnější informace o



oblačnosti a stavu atmosféry. Proto jsou družicová data v současnosti meteorology využívána k zpřesňování analýzy aktuálního stavu počasí, ať jsou již vyhodnocována subjektivně, nebo vstupují do numerických modelů počasí a vzhledem k jejich již téměř padesátileté dostupnosti mají značný význam i pro klimatologii (Řezáčová, 2007). Základní dělení meteorologických družic se provádí vzhledem k charakteru jejich oběžné dráhy na družice na (kvazi)polárních drahách (polární družice) a družice na geostacionární dráze (geostacionární družice). Kromě těchto lze dále vydělit méně častější družice např. ekvatoriální, experimentální, podrobnosti viz. M. Capderou (2005).

### 1.1.1 Polární družice

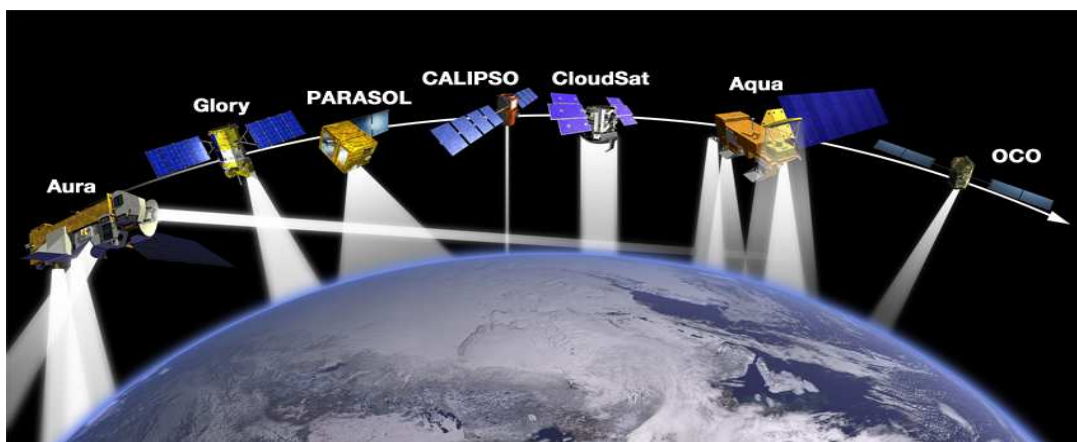
Tyto družice na nízkých orbitách obíhají ve výškách kolem 600 km – 1500 km po drahách přibližně kolmých na rovinu zemského rovníku, takže při každém přeletu přelétají polární oblasti, v kterých je tím pádem největší hustota měření těmito družicemi. Do této kategorie patří mimo jiné evropské družice METOP a americké NOAA/POES shodně využívající radiometr AVHRR (*Advanced Very High Resolution Radiometer*), jehož rozlišovací schopnost je v nadiru (neboli v tzv. poddružicovém bodě; pozorovateli v tomto bodě by družice prolétla přesně zenitem) 1,1 km × 1,1 km a 2,5 km × 4,5 km na okraji snímaného pásu. Informace o spektrálních pásmech snímaných v současné šestikanálové verzi radiometru jsou uvedeny v tabulce 1.1.

Pořadové číslo kanálu	Spektrální pásmo	Popis kanálu
kanál 1	0,58–0,68 μm	viditelné (červené) záření (VIS)
kanál 2	0,725–1,0 μm	blízké infračervené záření (NIR)
kanál 3A	1,58–1,64 μm	blízké infračervené záření (NIR)
kanál 3B	3,55–3,93 μm	(střední) infračervené záření (IR), atmosférické okno
kanál 4	10,3 – 11,3 μm	tepelné záření (IR), atmosférické okno
kanál 5	11,5 – 12,5 μm	tepelné záření (IR), atmosférické okno

Tab.1.1: Spektrální kanály přístroje AVHRR/3

Data se z polárních družic vysílají přímo k uživatelům hned po naměření. Jejich příjem je tedy možný jen v době přímé geometrické viditelnosti družice nebo z různých redistribučních zdrojů. Určité konkrétní území se vyskytuje na snímcích jedné družice obvykle dvakrát během 24 hodin, ovšem v polárních oblastech, kde dochází k výrazným překryvům sousedních přeletů, mohou být k dispozici mnohem častěji a pro studium těchto oblastí hrají polární družice nezastupitelnou roli.

Vzhledem k častému použití k validaci dat od družic operativně využívaných organizací EUMETSAT (podrobnosti v další podkapitole) je vhodné se zde zmínit ještě o výzkumných družicích NASA Terra a Aqua. Družice Terra má oběh načasovaný tak, aby protínala rovinu rovníku směrem od severu k jihu ráno, zatímco družice Aqua a satelity s ní spojené stejnou oběžnou drahou v tzv. A-train (viz obr. 1.2) procházejí skrz rovinu rovníku od jihu k severu odpoledne (proto zkratka A – *afternoon*) ([www.nasa.org](http://www.nasa.org)). Terra a Aqua nesou radiometr MODIS (*Moderate Resolution Imaging Spectroradiometr*), který má 36 kanálů, z toho 22 vhodných pro studium oblačnosti, a v závislosti na kanálu nabízí data v rozlišení v nadhlavníku 250 m pro 2 kanály, 500m pro 5 kanálů a pro ostatní 1 km (Řezáčová, 2007).



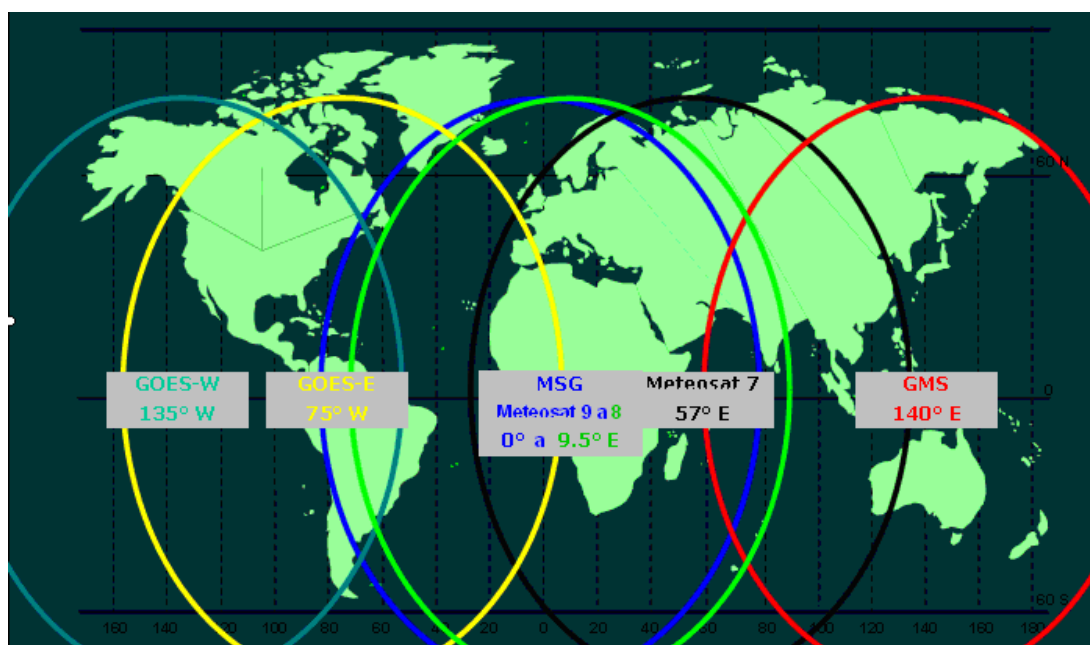
Obr. 1.2: Schematické znázornění družic tvořících tzv. A-train. (Převzato od NASA z <http://www-calipso.larc.nasa.gov/about/atrain.php>)

AVHRR i MODIS jsou oba pasivními senzory a uvedení tzv. A-train družic, zejména těch nesoucích aktivní senzory CloudSat (používající CPR – *Cloud Profiling Radar*) a CALIPSO (nesoucí mimo jiné přístroj CALIOP – *Cloud-Aerosol Lidar with Orthogonal Polarization*), velice zlepšilo možnosti validace dat, viz např. Karlsson (2010). Družice Terra a Aqua obsahují mimo jiné i 4 senzory

CERES (*the Clouds and the Earth's Radiant Energy System*), které jim umožňují měřit množství slunečního záření odráženého Zemí a množství Zemí emitovaného dlouhovlnného záření, což jsou důležité veličiny v globální radiční bilanci.

### 1.1.2 Geostacionární družice

Družice na geostacionární (geosynchronní) dráze, zkráceně geostacionární družice, obíhají po kruhových drahách o poloměru přibližně 42 168 km (geostacionární dráha se nachází přibližně ve výšce 35 786 km nad rovníkem) v rovině totožné s rovinou zemského rovníku proto, aby jejich oběžná rychlost postačila k dosažení rovnováhy mezi odstředivou a gravitační silou a zároveň jejich úhlová rychlost odpovídala úhlové rychlosti Země. Pro vyrovnání vlivů ostatních planet Sluneční soustavy a hlavně Měsíce musí být družice stabilizovány rotací nebo korekčními raketovými motory. Družice pak zdánlivě visí nad jedním určitým bodem na rovníku a jsou umísťovány tak, aby snímaly především zájmovou oblast (viz obr. 1.3) a v současnosti je jimi souvisle pokryt celý zemský povrch ohraničený přibližně 60. až 70. stupněm severní a jižní šířky, jak je vidět na obrázku 1.3. Ve vyšších zeměpisných šířkách již tyto družice nejsou moc dobře použitelné a limitem jejich viditelnosti je přibližně 81. rovnoběžka na obou polokoulích (Řezáčová a kol., 2007).



Obr. 1.3: Schematické umístění geostacionárních družic (stejnou barvou jako název a poledník je vyznačena přibližná oblast na severní polokouli snímána družicí).

Oproti polárním družicím jsou geostacionární družice více vzdáleny od zemského povrchu, což se odráží v jejich obecně menším rozlišení. Mají ovšem tu výhodu, že mohou kontinuálně (spíše v relativně krátkých intervalech odvislých od rychlosti přístrojů) snímat přivrácenou část Země. Nejdůležitějšími geostacionárními družicemi pro oblast Evropy jsou v současnosti bezesporu evropské družice Meteosat druhé generace (MSG) nesoucí mimo jiné přístroj GERB (*Geostationary Earth Radiation Budget*) a hlavně radiometr SEVIRI (*Spinning Enhanced Visible and Infrared Imager*). GERB snímá dopadající přímou sluneční radiaci, sluneční radiaci odraženou od TOA (*the top of the atmosphere* - horní hranice atmosféry) a tepelné záření emitované Zemí. Tedy veličiny, ze kterých se skládá radiální bilance Země. Přístroj SEVIRI snímá záření přicházející od Země ve 12 spektrálních kanálech, viz tab. 1.2, každých 15 minut a to v rozlišení 3 km × 3 km v nadiru a přibližně 4 km × 6 km v oblasti střední Evropy, kromě 12. kanálu, ten má rozlišení větší. Po získání jsou data nejdříve zkontrolována a korigována v řídicím centru organizace EUMETSAT a až poté distribuována uživatelům (Řezáčová,2007).

Pořadové číslo kanálu	Oficiální označení	Spektrální pásmo	Popis kanálu
01	VIS 0.6	0,56–0,71 μm	viditelné záření
02	VIS 0.8	0,74–0,88 μm	blízké infračervené záření
03	NIR 1.6	1,50–1,78 μm	blízké infračervené záření
04	IR 3.9	3,48–4,36 μm	(střední) infračervené záření, atmos. okno
05	WV 6.2	5,35–7,15 μm	pásmo absorpce vodní parou
06	WV 7.3	6,85–7,85 μm	pásmo absorpce vodní parou
07	IR 8.7	8,30–9,10 μm	tepelné záření, atmosférické okno
08	IR 9.7	9,38–9,94 μm	tepelné záření, pásmo absorpce O <sub>3</sub>
09	IR 10.8	9,80–11,80 μm	tepelné záření, atmosférické okno
10	IR 12.0	11,00–13,00 μm	tepelné záření, atmosférické okno
11	IR 13.4	12,40–14,40 μm	tepelné záření, pásmo absorpce CO <sub>2</sub>
12	HRV	0,50–0,90 μm	viditelné záření, vysoké rozlišení (1 km v nadiru)

Tab. 1.2: Spektrální kanály přístroje SEVIRI

## 1.2 EUMETSAT a síť „SAFů“

Organizace EUMETSAT (European Organisation for the Exploitation of Meteorological Satellites) byla utvořena osamostatněním od ESA (European Space Agency) roku 1986 a jejím sídlem se stal německý Darmstadt. V současnosti je EUMETSAT tvořen 26 členskými státy a dalších 5 států má status spolupracujícího státu. Česká republika se stala plným členem organizace v květnu roku 2010 po pěti letech členství jako spolupracující stát. EUMETSAT provozuje systém meteorologických satelitů monitorujících Zemi, které nepřetržitě dodávají data, snímky a produkty informující o klimatu a stavu počasí, a těmi pak zásobuje národní meteorologické organizace.

Cílem samotné organizace, jak se můžeme dočíst na stránkách <http://www.eumetsat.int/Home/Main/AboutEUMETSAT/Strategy>, je maximalizovat svou užitečnost pro členské a spolupracující státy, hrát vůdčí roli v evropských programech distančního pozorování a být důvěryhodným celosvětovým partnerem v zajišťování a poskytování dat z geostacionárních a polárních družic. Pro účely operativní předpovědi počasí EUMETSAT reaguje na požadavky evropských národních meteorologických služeb ohledně družicových dat pro podporu nowcastingu, krátkodobé a dlouhodobé předpovědi. To vyžaduje pozorování nejen stavu atmosféry, ale i zemského povrchu, oceánu a kryosféry.

Pro účely monitorování klimatu, jakožto úlohy vyžadující celosvětovou spolupráci, se EUMETSAT začleňuje do různých světových organizací. Reaguje například na požadavky vyjádřené operativním systémem GCOS (Global Climate Observing System), který byl založen r. 1992 frakcí organizace OSN UNFCCC (United Nations Framework Convention on Climate Change) a je spolufinancován organizacemi jako Světová meteorologická organizace (WMO), IOC (Intergovernmental Oceanographic Commission) organizace UNESCO, programem organizace OSN pro životní prostředí UNEP a organizací ICSU (International Council for Science). Pro naplňování požadavků GCOS na distanční pozorování koordinuje EUMETSAT své klimatologické aktivity v rámci skupin CEOS (Committee on Earth Observation Satellites) a GEO (Group on Earth Observations) dále generuje tzv. FCDR (Fundamental Climate Data Records) a TCDR (Thematic Climate Data Records) a zajišťuje síť tzv. SAFů (od slova Satellite Application Facility), to jest center pro práci s úzce specializovanými i daty.

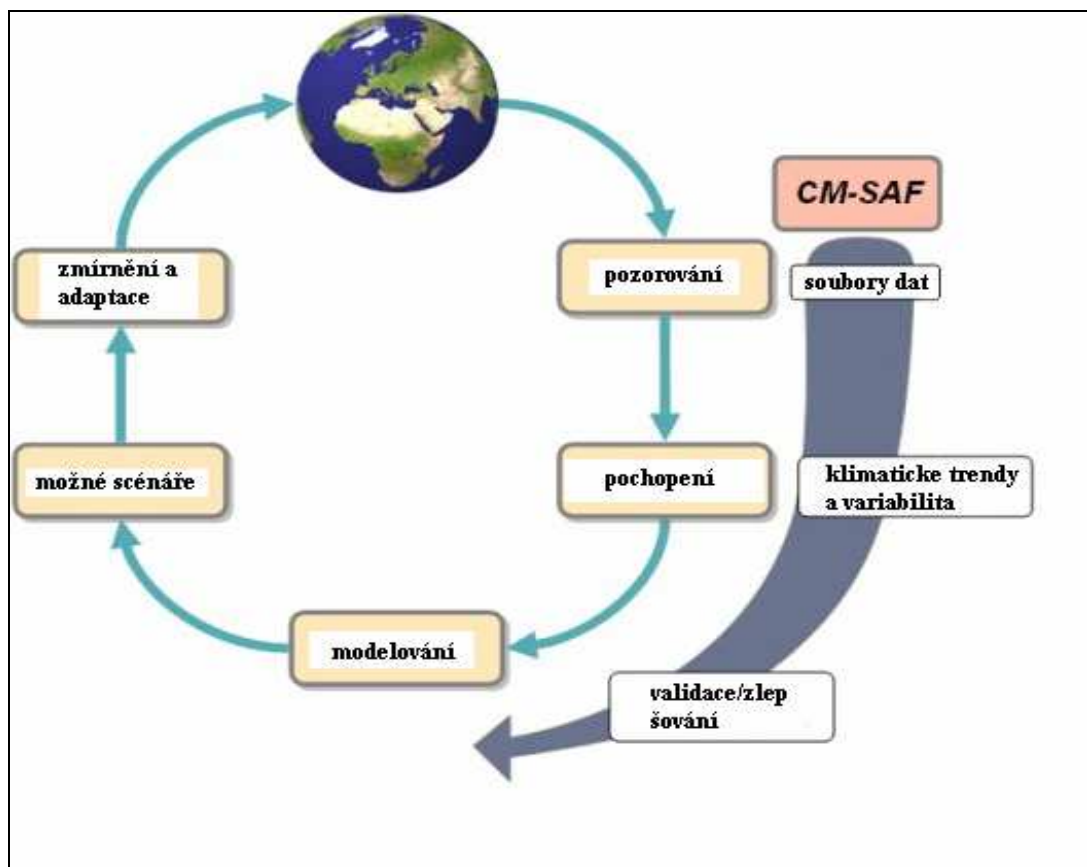
Pod pojmem FCDR se rozumí přímo satelitem měřené veličiny jako např. radiance a jasová teplota, které poté prochází fázemi takzvaného „reprocessingu“, kalibrace a korekce. Z nich se potom vyvozují data skrývající se pod zkratkou TCDR. Jsou to například různé informace o teplotě, vlhkosti, množství ozonu a vegetaci. Vedle toho SAFy jsou specializovanými odbornými centry, která jsou zaměřená na zpracování dat z družic. Koncept sítě SAFů byl přijat roku 1992 v Darmstadtu a v současné době jich existuje osm s různým specifickým zaměřením. Každý SAF je řízen národní meteorologickou službou vybraného člena organizace EUMETSAT. To ovšem neznamená, že by se na něm ostatní členové a spolupracující státy nemohli spolupodílet. Vše je pak dozíráno a kooperováno EUMETSATem.

Co se týče jednotlivých operativně fungujících SAFů, jsou to OSI-SAF (SAF on Ocean and Sea Ice) zaměřený na monitorování oceánu a mořského ledu, NWP-SAF (SAF on Numerical Weather Prediction) cílený na vývoj softwaru na přípravu družicových dat pro účely numerické předpovědi, NWC-SAF (SAF on Support to Nowcasting and Very Short Range Forecasting) zabývající se vytvářením produktů z družicových dat vhodných pro nowcasting a velmi krátkodobou předpověď, LSA-SAF (SAF on Land Surface Analysis), jehož hlavním úkolem je zkoumat možnosti využití družicových dat k monitorování zemského povrchu, O3M-SAF (SAF on Ozone & Atmospheric Chemistry Monitoring) založený pro monitorování ozónu a dalších stopových plynů, aerosolů a UV záření pomocí družic, GRAS-SAF (SAF on GRAS Meteorology), v jehož rámci jsou vyvíjeny algoritmy pro zpracovávání dat z přístrojů GRAS (Global Navigation Satellite System Receiver for Atmospheric Sounding), což je GPS přijímač nesený družicí METOP schopný pomocí speciální metody proměřit až 500 vertikálních profilů teploty, vlhkosti a tlaku denně), H-SAF (SAF on Support to Operational Hydrology and Water Management) založený pro vývoj a distribuci dat a produktů používaných v hydrologii a konečně CM-SAF (SAF on Climate Monitoring), který podrobněji rozebereme v následující podkapitole a jehož produkty v této práci využíváme.

### **1.3 CM-SAF**

Tento projekt byl vytvořen německou meteorologickou službou DWD (Deutscher Wetterdienst) za přispění finské, belgické, nizozemské, švédské a švýcarské národní meteorologické služby. Jeho počáteční pětiletá fáze mající za cíl

vytvoření infrastruktury byla odstartována roku 1999. V roce 2004 začala tzv. IOP (Initial Operations Phase) nahrazená roku 2007 fází CDOP (Continuous Development and Operations Phase), která byla v březnu 2012 nahrazena druhou CDOP. Jednotlivé fáze na sebe navazují a postupem času dochází k rozšiřování a zkvalitňování služeb a produktů CM-SAFu. Obecně lze říci, že jsou v rámci organizace generovány a archivovány vysoce kvalitní datové soubory použitelné pro studium klimatu a pomocí nejmodernějších algoritmů se tu z družicových měření získávají informace o klimatických proměnných zemského systému, přesně dle již výše zmiňovaných nadnárodních požadavků. Schematický náčrt studia klimatu a jeho změny a zamýšlené přispění CM-SAFu v tomto procesu je znázorněn na obrázku 1.4.



*Obr. 1.4: Přispění CM-SAFu ke studiu klimatu (pro pozorování náležitými datovými soubory; pro porozumění klimatu podporou jeho studia, studia klimatických trendů a analýzami variability; ke zlepšování klimatických modelů dodáváním dat pro validaci).*

Pro účely CM-SAFu jsou používána data z polárních družic METOP a NOAA, na kterých se podílejí hlavně přístroje AVHRR, HIRS (High resolution Infrared Radiation Sounder), IASI (Infrared Atmospheric Sounding Interferometer), AMSU (Advanced Microwave Sounding Unit) a MHS (Microwave Humidity Sounder). Dále jsou používána data z radiometrů geostacionárních družic MSG - SEVIRI a GERB. Pro podporu měření radičních toků na TOA (horní hranici atmosféry) se využívají data ze systému CERES (Clouds and Earth's Radiant Energy System) neseného družicemi TERRA a AQUA a dále, pro zajištění konzistentních časových řad celkového množství vodní páry (total column water vapor) nad oceány v časovém rozpětí let 1987–2005, se používají data ze senzorů série SSM/I (Special Sensor Microwave/ Imager) nesených polárními družicemi série DMSP (Defense Meteorological Satellites Program).

Samotná tvorba struktury CM-SAFu probíhala následovně. Během IOP byly vytvořeny operativní procedury pro rychlé zpracování velkého množství dat a produkty byly dostupné v téměř reálném čase. Postupy získávání dat a tvorby produktů se však s časem měnily a navíc radiance používané jako vstup byly jen formálně kalibrovány, nebyly tedy očištěny od chyb spjatých se změnou senzorů a od dalších s přístroji spojených chyb.

Na data lépe vyhovující potřebám studia klimatu byla zaměřena tzv. CDOP fáze, kdy byly vytvářeny dlouhé homogenní časové řady produktů CM-SAFu. Současně dostupné produkty můžeme tedy rozdělit do dvou kategorií, běžně a operativně produkované sady dat a datové soubory založené na citlivé kalibraci. Produkty první kategorie spadají pod již zmíněný pojem CDR (Climate Data Records) a jsou dány časovou integrací okamžitých odhadů geofyzikálních veličin pro získání denních a měsíčních průměrů. Tato data jsou vhodná pro studium denního, maximálně sezónního měřítka, např. monitorování extrémních jevů, ale jsou dostatečně přesná pro využití např. v oblasti solární energetiky.

Pro možnost použití ve větších časových měřítkách by byla třeba lepší kalibrace senzorů než jen automatická, která je používána pro tento druh dat. Proto existuje druhá skupina dat nazývaná, „tematická“ CDR, TCDR. Jedná se o CDR data prošlé reprocessingem za předpokladu odborné úpravy o korekce přístrojových chyb (zhoršující se senzory, změna oběžné dráhy atd.) a „získávacích“ (změny algoritmů atp.) chyb. Takové produkty jsou použitelné pro studium procesů časových měřítek od denních po meziroční. Pro větší měřítka a např. pro studium klimatické variability



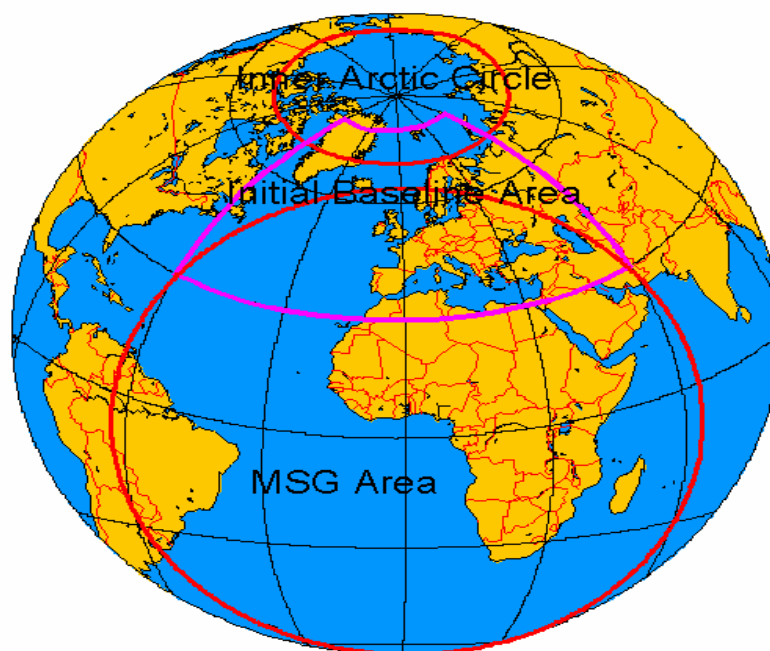
by bylo potřeba absolutně kalibrovaných a expertně očištěných dat, aby mohla být objevena i sebemenší změna v chování systému. Je zapotřebí i dostatečná (alespoň 30letá) délka takových řad. Toto bude splňovat budoucí třetí kategorie produktů, která by měla být k dispozici na konci fáze CDOP. Jednotlivé operativně používané produkty s upřesňujícími informacemi jsou uvedeny v tabulce 1.3. Tabulka má pouze shrnující charakter, produkty používané v této práci budou dále podrobněji rozebrány a bude osvětlen jejich význam.

Typ produktu	Vlastní produkty	Senzor	Oblast dostupnosti	Dostupné od:	Rozlišení (prostorové a časové)
Oblačné produkty	CFC (Cloud Fractional Cover)	AVHRR AVHRR SEVIRI	Baseline Arktida oblast snímaná Meteosatem	01.11.2004 01.01.2009 01.09.2005	(15 km) <sup>2</sup> M,D a M, D, MDC pro SEVIRI
	CTY (Cloud Type)				
	CTP (Cloud Top Pressure)				
	CTH (Cloud Top Height)				
	CTT (Cloud Top Temperature)				
	CPH (Cloud Phase)				
	COT (Cloud Optical Thickness)				
	CWP (Cloud Water Path)				
Vlhkostní produkty	HTW (Total precipitable water)	ATOVS	Globální pokrytí	01.01.2004	(90 km) <sup>2</sup> M, D
	HLW (Layered precipitable water)				
	HSW				
Záření na Zemském povrchu	SIS (Incoming short-wave radiation)	AVHRR AVHRR SEVIRI	Baseline Arktida (jen pro SAL) Oblast Meteosatu	01.11.2004 01.01.2009 01.09.2005	(15 km) <sup>2</sup> M, D SEVIRI i MDC
	SAL (Surface albedo)				
	SNS (Net short-wave radiation)				
	SNL (Net long-wave radiation)				
	SDL (Downward long-wave radiation)				
	SOL (Outgoing long-wave radiation)				
	SRB (Surface radiation budget)				
TET (Emitted thermal radiative flux)	TIS (Incoming solar radiative flux)	GERB + CERES (merged dataset)	Oblast Meteosatu Baseline	01.02.2004	(45 km) <sup>2</sup> M, D, MDC
	TRS (Reflected solar radiative flux)				

*Tabulka 1.3: Operativní (tzn. pravidelně aktualizované) produkty (M značí měsíční průměry, D denní průměry, MDC měsíční průměr denního chodu. Oblasti dostupnosti viz obr. 1.5).*

## 1.4 Oblačné produkty

Už bylo řečeno, že oblačnost, jakožto dominantní prvek planetárního albeda, hraje důležitou roli v koloběhu vody a globálním energickém cyklu. Tím pádem je pro monitorování klimatu nezbytné sledovat změny množství celkové oblačnosti, ale nejen to, důležité je i studium potenciálních zpětných vazeb směřujících například ke změnám teplot a obsahu vodní páry v atmosféře. Potřebná je také informace o typu oblačnosti a to hlavně kvůli tomu, že rozličné typy oblaků mají různé odrazivostní, transmisivní a absorpční charakteristiky. Tato motivace vedla CM- SAF k vytvoření oblačných produktů, jejichž využití (ať už operativních nebo retrospektivně produkovaných) můžeme vidět kupříkladu v evaluaci modelování oblačnosti předpovědními i klimatickými modely, jak ukázali např. Karlsson a kol. (2008), Roebeling a kol. (2009) a Willén (2008).



Obr.1.5: Oblasti dostupnosti produktů (převzat z manuálu cloud products).

IAC (Inner Arctic Circle), IBA (Initial Baseline Area).

Na gridboxu velikosti  $15\text{km} \times 15\text{ km}$  je definováno 6 operativních produktů, každý dvou různých druhů. Jeden typ vychází z dat z polárních družic (NOAA/METOP, AVHRR) a druhý z geostacionárních dat (METEOSAT, SEVIRI). Oba druhy nabízejí data jak v denních, tak měsíčních průměrech, pouze

geostacionární typ však existuje i ve formě měsíčního průměru denního chodu. Ohledně zpracování dat a tvorby produktů stojí za oběma typy odlišné algoritmy, které se navíc s časem měnily, což vyústilo v existenci různých vývojových verzí, a také se objevují určité rozdíly mezi geografickými oblastmi. Zatímco produkty v IAC (viz obr. 1.5) jsou udávány v Lambertově projekci, produkty v oblasti IBA a MSG disku (viz obr. 1.5) jsou v sinusoidální projekci. Zájemce o podrobnější informace o způsobech zpracování a rozdílech mezi oblastmi bychom odkázali na manuál od Karlssona a kol. (2011). V této práci budou počítána data z oblasti České republiky, která spadá do IBA. Nyní si podrobněji představíme námi analyzované operativní oblačné produkty.

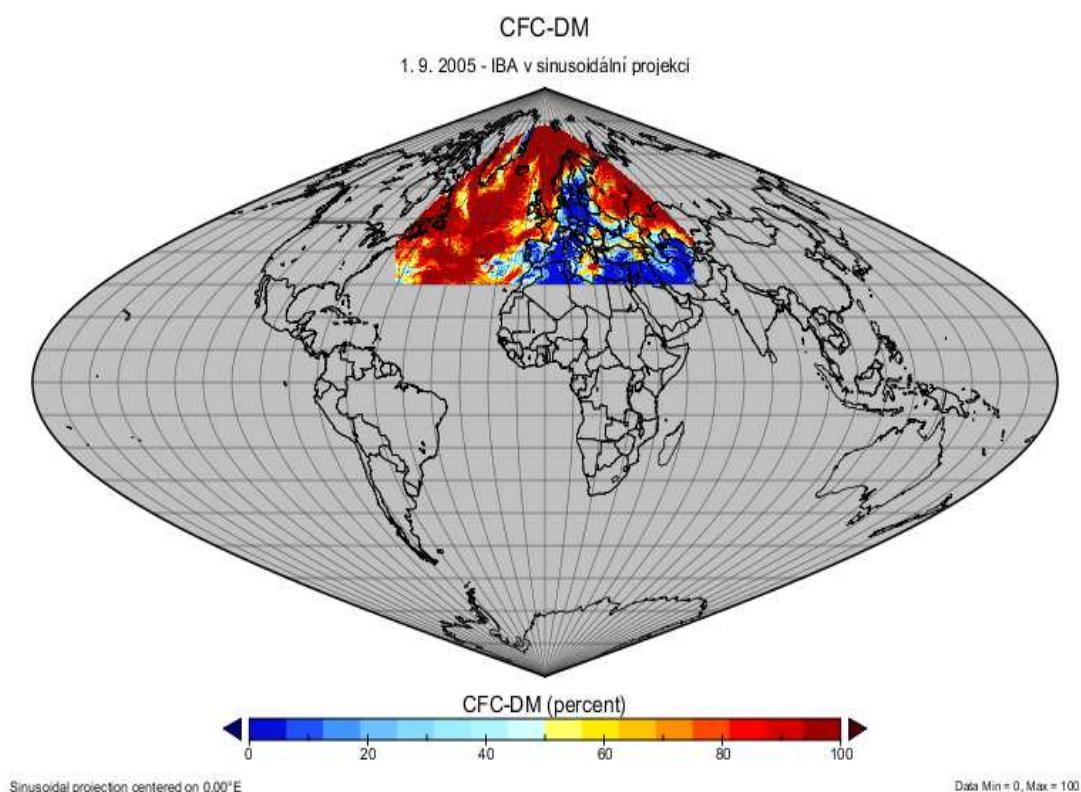
#### **1.4.1 CFC (Cloud Fractional Cover)**

CFC, procentuální pokrytí oblačností, je počítáno přímo pomocí algoritmů NWC-SAFu pro výpočet zakrytí oblačností (CMA-Cloud Mask). Algoritmy pro polární družice a MSG se liší a jsou detailně představeny ve článcích Dybbroe et al. (2005a a 2005b) a Derrien a Leglau (2005). Oba algoritmy jsou založeny na multispektrální prahové technice aplikované na každý pixel družicového obrazu, přičemž se využívá typických odlišností vzhledu oblačnosti v různých spektrálních pásmech. K tomu, aby byl pixel označen jako oblačný nebo bez oblačnosti, musí být splněno mnoho prahových testů. V generování prahů jsou odlišné přístupy mezi algoritmy pro zpracování dat ze SEVIRI a z AVHRR. Zatímco algoritmy pro data z MSG využívají klimatologické odrazivosti povrchu a povrchové teploty oceánu, AVHRR algoritmy využívají produkty asimilace z numerických modelů předpovědi počasí. Pro každý družicový snímek a přelet je generována unikátní sada prahových podmínek.

Finální produkt CFC je počítán jako podíl pixelů družicového snímku plně nebo částečně pokrytých oblačností oproti všem pixelům nacházejícím se v gridboxu  $15 \text{ km} \times 15 \text{ km}$  v sinusoidální projekci mapy a je uváděn v jednotkách procent. Měsíční průměr CFC je počítán z příslušných denních průměrů CFC, přičemž je nutné, aby v daném měsíci bylo k dispozici alespoň 20 denních průměrů. Pro výpočet denního průměru je potřeba přinejmenším 6 přeletů či termínů měření MSG nad daným gridboxem. Pro výpočet měsíčního průměru denního chodu je pro každý termín spočten měsíční průměr ze všech dnů, ze všech termínů v jednom dni je pak

složen denní chod. Každý termín musí být k dispozici v alespoň dvaceti dnech měsíce, aby bylo možno vypočítat jeho průměr.

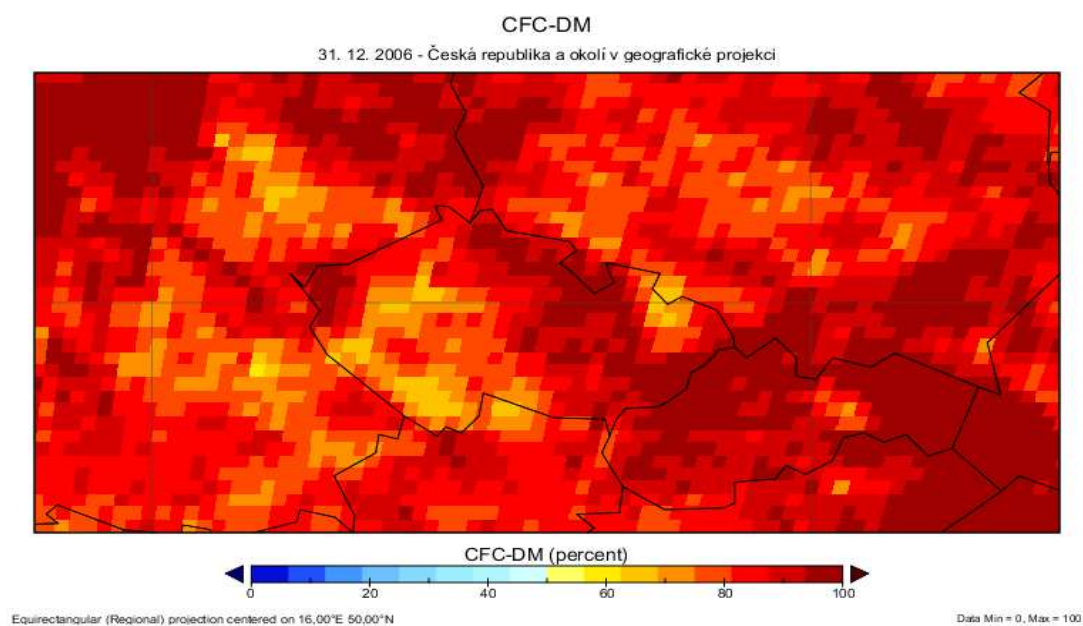
Všechny produkty musí splňovat požadavky na přesnost definované v Service Specification Document. Nejistoty vydávaných produktů jsou pak každý rok zkoumány a konfrontovány s těmito požadavky a výsledky jsou shrnuty v CM-SAF Annual Validation Report (nejaktuálnější z r. 2010). Validace se provádí většinou vzhledem k pozemním měřením SYNOP a družicovým produktům MODIS. Pro produkt CFC jsou požadavky na absolutní chybu následující. Pro AVHRR může být střední chyba v oblasti IBA  $\pm 10\%$ , přičemž větší odchylky se mohou objevit nad oceánem, v oblasti IAC  $\pm 30\%$ . Větší odchylky od skutečnosti se mohou vyskytovat během velmi studených měsíců polární zimy. Produkt založený na datech ze senzoru MSG SEVIRI je dodáván s přesností  $\pm 10\%$ , ovšem nad oceány a zejména na okrajích MSG disku může docházet k většímu procentuálnímu nadhodnocování oblačnosti a nad tropickými oblastmi naopak k podhodnocování.



*Obr. 1.6: Příklad denního průměru CFC od MSG nad Initial Baseline Area z 1. 9. 2005. Zpracováno programem Panoply.*

Známými omezeními přesnosti produktů, jak jsou vyjmenovány v Karlsson a kol. (2011), jsou dále: *rozdíly zapříčiněné odlišnými pozorovacími úhly*, kdy produkt

získaný z MSG udává oproti AVHRR tím větší sumy oblačnosti, čím je daná oblast severněji. *Efekt subpixelové oblačnosti*, kdy jsou pixely družicového snímku označeny jako oblačné, i když nejsou pokryty oblačností cele. Tento jev vede k nadhodnocování pokrytí oblačnosti a to výrazněji nad oceány, přičemž současné výsledky validací ukazují, že tím více trpí produkty SEVIRI. *Rozdíly mezi pevninou a oceánem*, jelikož, kromě efektů nežádoucích odrazů a zrcadlení Slunce na vodní hladině, je nad oceány obecně lepší kontrast mezi oblačnými a bezoblačnými oblastmi. *Falešná umělá oblačnost podél pobřeží* kvůli nepřesné lokalizaci ostré hranice voda a pevnina a používání neplatných prahů. Dvě poslední omezení se České republiky pochopitelně netýkají, stejně tak jako rušivé horizontální a vertikální čáry v starších verzích produktu v oblastech vzdálenějších od hlavní přijímací stanice Offenbach.



*Obr. 1.7: Příklad denního průměru CFC získaného z MSG nad Českou republikou. Zpracováno programem Panoply.*

Zato vliv *chybějící nízké oblačnosti za soumraku*, kdy při polohách Slunce nízko nad obzorem může docházet k podhodnocování kolem 10% pokrytí oblačností, a *denní a noční rozdíly*, které jsou zapříčiněné omezenými možnostmi použití senzorů v noci, a *vliv extrémně studených povrchů během polární zimy*, který se může projevit i na našem území při extrémně mrazové situaci, mohou vést ke zkreslení výsledného CFC produktu i nad Českou republikou. Omezení přesnosti

produktů v nočních hodinách nebylo dosud kvantifikováno, jelikož pro validaci potřebná referenční synoptická pozorování jsou v noci taktéž limitována (Karlsson a kol., 2011).

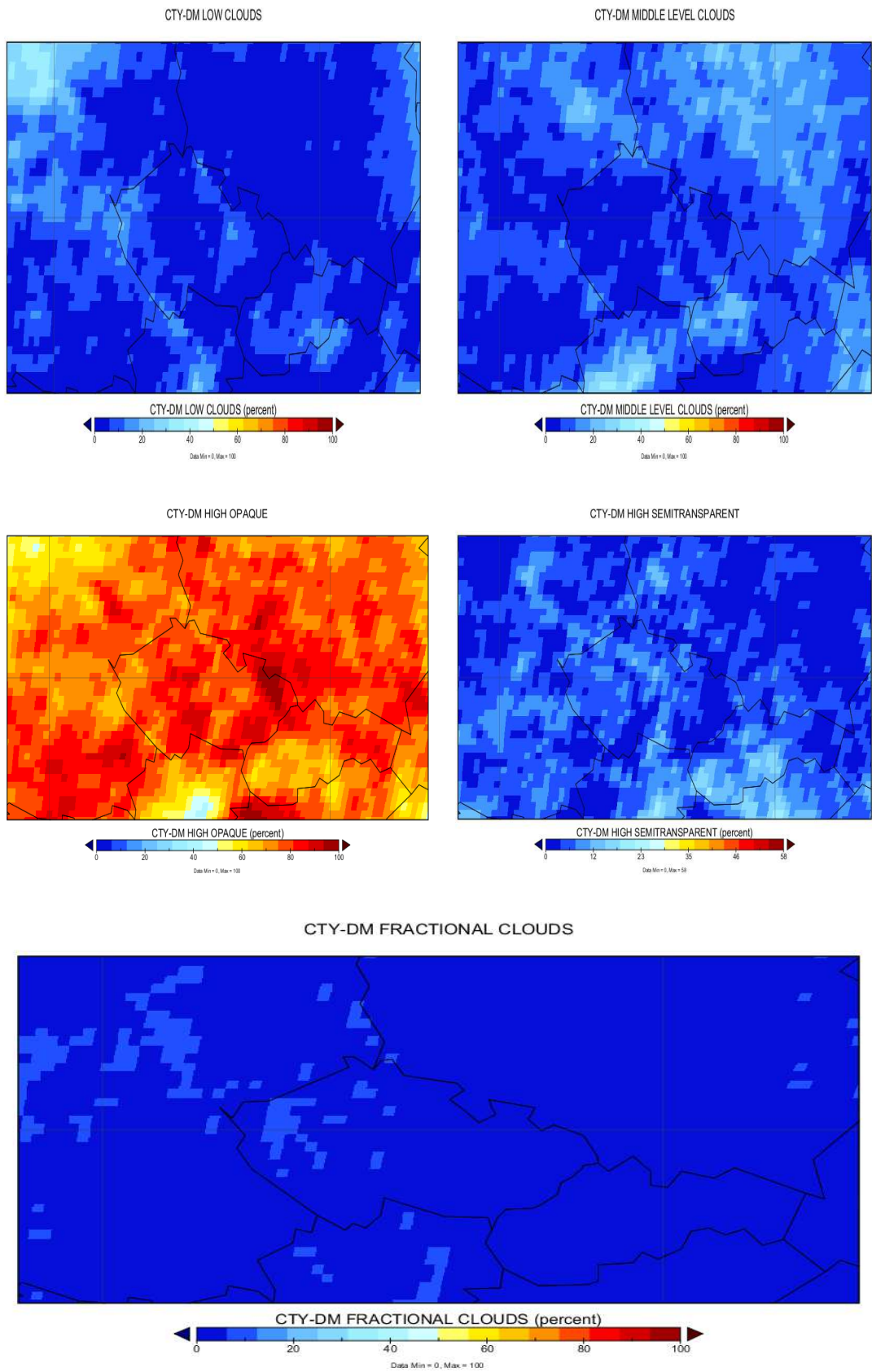
#### **1.4.2 CTY (Cloud Type)**

CTY, typ oblačnosti, je produkt vycházející z algoritmů Cloud Type NWC-SAFu, ale je méně detailní a původních dvacet kategorií typů oblačnosti shrnuje do následujících pěti kategorií: nízké oblaky (Low clouds), oblaky středního patra (Middle level clouds), vysoké neprůhledné oblaky (High opaque clouds), vysoké polopropustné oblaky (High semitransparent clouds) a podpixelová, útržkovitá oblačnost (Fractional clouds). Každý pixel, který je za použití již zmiňovaných prahových algoritmů označený na satelitním snímku jako oblačný, je zařazen do jedné z výše jmenovaných kategorií a poté je počítáno procentuelní zastoupení oblačných pixelů jedné kategorie vůči všem oblačným pixelům v daném gridboxu. Produkt tedy sestává z pěti separátních částí, které dohromady musejí dávat 100%. Pro vyjádření absolutního procentuelního příspěví dané kategorie k pokrytí gridboxu oblačností je třeba vynásobit její procentuelní výskyt korespondující hodnotou CFC.

Všechny pixely označené na satelitním snímku jako oblačné jsou zařazeny do jedné z těchto kategorií. Základním principem je vertikální rozdělení oblaků pomocí jasových teplot, dále se identifikují oblaky tvořené ledovými částěčkami využitím různých charakteristik odrazivosti ledu a vody v NIR a středním IR pásmu a polopropustné oblaky pomocí rozdílných transmisivit v atmosférických oknech v NIR a IR pásmu. Podrobnější popis metod detekce a rozlišení různých typů oblačnosti je velice srozumitelnou formou popsán v Řezáčové (2007).

Po identifikaci typu oblačnosti na pixelu je počítáno procentuelní zastoupení oblačných pixelů jedné kategorie vůči všem oblačným pixelům v daném gridboxu. Průměrování je prováděno dle stejných pravidel jako u produktu CFC a je zřejmé, že i již vyjmenovaná omezení přesnosti produktu CFC budou platná i pro CTY. Schopnost rozlišit jednotlivé druhy oblaků je však mnohem citlivější na radiačních poměrech sledované oblasti, geometrii zorného pole atd. To vše naznačuje, že kvalita produktu by se mohla lišit v závislosti na sezóně, což bude také předmětem našeho studia.





Obr. 1.7: Denní průměr produktu CTY pro Českou republiku a okolí z 28. 2. 2007 udávající relativní příspěvek pěti typů oblačnosti k celkové oblačnosti.



## **2 Analyzovaná data a použité statistické metody**

V této kapitole bude podán názorný přehled zdrojů a metodiky autorského výzkumu nezbytný pro porozumění výsledkům další kapitoly. Navíc již samotná konstrukce námi zkoumané úlohy je výsledkem netriviální práce a ještě před počátkem hlubších analýz navozuje zajímavé otázky k diskuzi.

### **2.1 Soubory dat a koncept validace**

#### **2.1.1 Oblačné produkty družicových dat**

Produkty vybrané pro analýzu jsou (v sekci 1.2 již představené) operativní produkty organizace CM-SAF a to denní produkty CFC a CTY. Studovány byly tedy denní průměry veličin množství a typu oblačnosti odvozené z dat radiometru MSG – SEVIRI. Kvůli velkému množství dodávaných dat, nemohou být družicové informace ukládány ve formátech typu např. ASCII, proto CM-SAF poskytuje své produkty ve vysoce specializovaných binárních datových formátech vhodných pro ukládání a řízení takových multidimenzioálních datových souborů. Práce s velkými soubory dat se díky těmto formátům ulehčí, ovšem za cenu použití specializovaného softwaru. CM-SAF distribuuje své operativní produkty ve formátech HDF5 (Hierarchical Data Format) a NetCDF (Network Common Data Form), což byl původní formát i námi analyzovaných dat, jejichž podrobnější popis je v tabulce 2.1.

Celkové období pokryté produkty popsány v této tabulce je od 1. 9. 2005 do 25. 3. 2012. Předmětem analýzy však bude kvůli dostupnosti dat ze stanic jen období do 31. 12. 2011. V oblasti překrytu dat verze 210 a 300 byla použita data z verze 210. Celkově se jedná o 2313 dnů, ovšem po odečtení dnů bez dostupných družicových produktů nám k analýze zůstává 2283 denních produktů CFC a stejný počet denních CTY ve formátu NetCDF. Jak se liší různé verze produktů uvedené v tabulce 2.1 zde nebudeme blíže popisovat, jelikož na naši práci tyto změny nemají vliv, případně zájemce odkážeme na manuál Karlsson a kol. (2011). Za pozornost v tabulce 2.1 stojí rychlejší nahrazení verze CTY 320 verzí 330, než je tomu u CFC.

Soubory dat obou produktů byly dále zpracovávány za účelem možnosti porovnání s přízemními měřeními, čemuž bude věnována sekce přicházející na řadu po krátkém představení použitých referenčních dat.

<b>Verze</b>	210	300 , 310, 320	300, 310, 320, 330
<b>Pokryté období</b>	01.09.2005 - 09.05.2007	01.05.2007 - 30.04.2008 01.05.2008 - 31.12.2008 01.01.2009 - 25.03.2012	01.05.2007 - 30.04.2008 01.05.2008 - 31.12.2008 01.01.2009 - 31.12.2009 01.01.2010 - 25.03.2012
<b>Chybějící dny</b>	08.11.2005 24.09.2006 - 09.10.2006	14.05.2008 - 15.05.2008 18.04.2009 - 22.04.2009 15.08.2009 - 20.08.2009	14.05.2008 - 15.05.2008 18.04.2009 - 22.04.2009 15.08.2009 - 20.08.2009
<b>Skupina produktů</b>	operativní produkty	operativní produkty	operativní produkty
<b>Název</b>	CFC – Fractional Cloud Cover CTY - Cloud type	CFC – Fractional Cloud Cover CTY - Cloud type	CTY - Cloud type
<b>Oblast</b>	CM SAF baseline area (30N-80N, 60W-60E)	MSG full disk (včetně Evropy, Afriky, Atlantického Oceánu)	MSG full disk (včetně Evropy, Afriky, Atlantického Oceánu)
<b>Časové rozlišení</b>	denní	denní	denní
<b>Statistika</b>	průměr	průměr	průměr
<b>Prostorové rozlišení</b>	sinusoidální projekce (15x15 km <sup>2</sup> )	sinusoidální projekce (15x15 km <sup>2</sup> )	sinusoidální projekce (15x15 km <sup>2</sup> )
<b>Zdroj dat</b>	MSG (SEVIRI)	MSG (SEVIRI)	MSG (SEVIRI)

*Tabulka 2.1: Analyzované družicové produkty.*

### 2.1.2 Referenční staniční měření

Od ČHMÚ byla laskavě poskytnuta měření pokrývající požadované období pocházející z devíti meteorologických stanic rozprostřených na území České republiky (viz tabulka 2.2). Jednalo se o ASCII soubor části zprávy SYNOP. Zahrnuta jsou jen měření prováděná pozorovatelem, automatická měření nejsou začleněna.

Pozorování byla prováděna v každou celou hodinu, kromě stanic Kuchařovice, Tušimice, kde byla měření prováděna jen od 06:00 do 21:00. Na omezený denní režim pozorování přešly v průběhu zkoumaného období i stanice Kocelovice a Liberec. V záznamech prakticky ze všech stanic se však nezdá objevují přerušování a větší mezery mezi pozorováními zejména v nočních hodinách, takže k dispozici byl nepravidelný výběrový soubor. Na většině stanic vyjma Pardubic byla měření prováděna na více geograficky nepatrně odlišných místech současně, nebo jednomu staničnímu koordinátu náleželo více současných měření. Tato měření byla aritmeticky zprůměrována do jednoho údaje, s kterým bylo dále pracováno. V tabulce 2.2 jsou pak uvedeny jen zaokrouhlené staniční koordináty, což vzhledem k povaze výzkumu je nepřesnost hodná zanedbání.

Ze zprávy SYNOP byla použita data o hodnotách veličin  $N$ ,  $C_l$ ,  $C_m$  a  $C_h$ .  $N$  je pokrytí oblohy oblaky udávané v osminách celkové plochy,  $C_l$  označuje čistě jen výskyt některého z oblaků  $St$ ,  $Sc$ ,  $Cu$ ,  $Cb$ , podobně  $C_m$  a  $C_h$  uvádějí výskyt oblaků  $Ac$ ,  $As$ ,  $Ns$  a  $Ci$ ,  $Cc$ ,  $Cs$ . Význam těchto veličin budeme ještě na některých místech této práce diskutovat, pravidla kódování je možné získat přímo ze zprávy SYNOP ([http://old.chmi.cz/meteo/olm/Let\\_met/stan\\_krouzek/data/synop.pdf](http://old.chmi.cz/meteo/olm/Let_met/stan_krouzek/data/synop.pdf)). Zkratky oblaků se přidržují standardních označení z Mezinárodního atlasu oblaků.

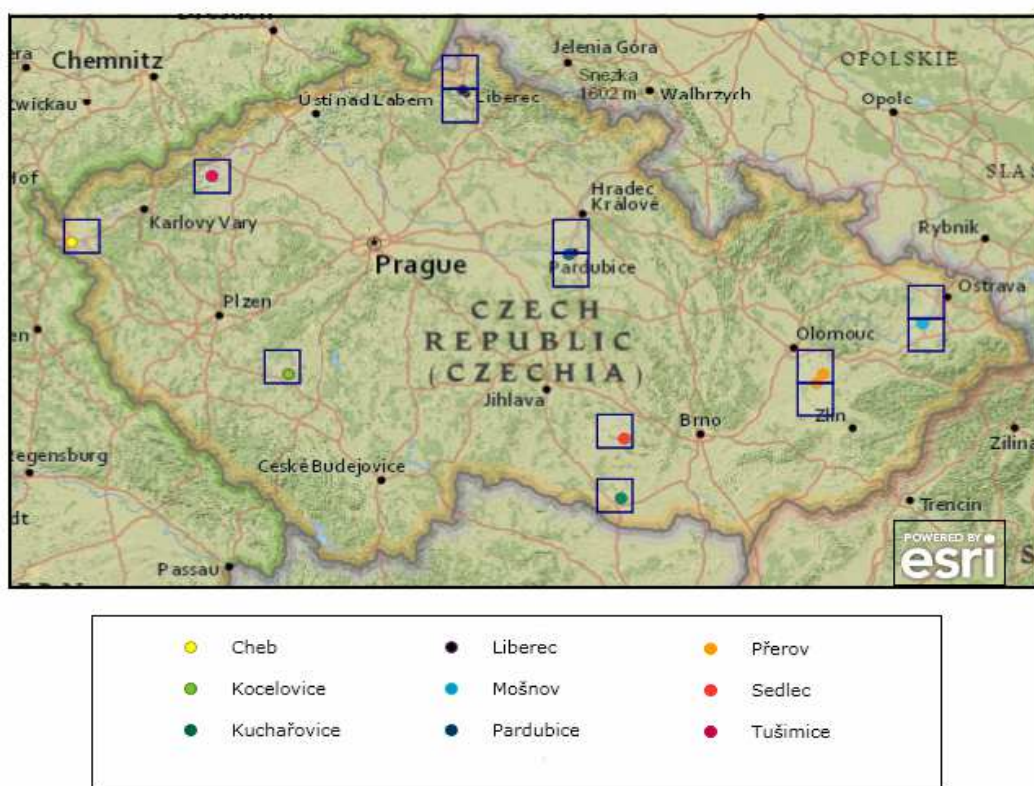
EG_GH_ID (identifikátor)	Název stanice	Zeměpisná šířka	Zeměpisná délka	Nadmořská výška [m]
L3CHEB01	Cheb	50,07°	12,39°	483
C1KOCE01	Kocelovice	49,46°	13,85°	508
B2KUCH01	Kuchařovice	48,88°	16,09°	334
U2LIBC01	Liberec	50,77°	15,03°	398
O1MOSN01	Mošnov	49,70°	18,12°	250
H3PARD01	Pardubice	50,02°	15,74°	225
O3PRER01	Přerov	49,45°	17,45°	210
B2SEDC01	Sedlec	49,17°	16,12°	474
U1KATU01	Tušimice	50,38°	13,32°	322

Tab. 2.2: Geografické údaje o využívaných stanicích.

### 2.1.3 Příprava dat

Jak poukazuje např. Reuter a kol. (2009), hlavním limitujícím faktorem validace satelitních dat je dostupnost nezávislých měření stejného parametru vhodně prostorově a časově rozložených. V této práci jsme jako referenční data použili

měření ze synoptických stanic blíže popsanych v předešlé podkapitole. Pro zajištění přibližně si odpovídajícího prostorového rozložení družicových dat s referenčními pozemními, bylo nejprve nutné vybrat pomocí sestaveného algoritmu z denního družicového produktu hodnotu náležející gridboxu, který je vzhledem ke geografické poloze stanici nejbližší. Polohy stanic a boxů, do nichž spadají, jsou vykresleny na obr. 2.1. Pokud stanice leží v okrajové oblasti gridboxu, byl s jejími daty porovnáván průměr hodnoty vyšetřovaného produktu ze dvou nejbližších boxů. Zkušebně byla oblast Chebu pokryta čtyřmi gridboxy ze všech stran, ale shodnost družicových dat s referenčními to ovlivnilo spíše negativně.



Obr. 2.1: Orientační ilustrace polohy stanic a k nim nejbližších gridboxů. Vytvořeno v ArcGIS Explorer.

Je potřeba si uvědomit, že v podobě N a CFC porovnáváme fyzikálně rozdílné veličiny, neboť pozorovatel nahlíží oblačnost zdola, vidí její spodní hranici a obsáhne svým měřením celý obzor, jehož velikost však silně závisí na vyskytující se oblačnosti. Naproti tomu MSG pozoruje střední Evropu pod nezanedbatelným úhlem, detekuje hlavně horní hranici a oblačnost musí být do gridboxu přeprojektována také v závislosti na její výšce v atmosféře. Tento problém bude ještě

dále v práci diskutován, nicméně volba gridboxu odpovídajícího stanici je velmi problematická.

Dále bylo potřeba k denním družicovým produktům sestavit denní průměry veličin přízemně měřených. Tento úkol byl algoritmicky obtížný, neboť měření sice byla prováděna v termínech zprávy, avšak, jak již bylo diskutováno, měření byla občas nekompletní či nepravidelná. Navíc je třeba přeskočit termíny, v nichž bylo měření ve zprávě sice zaznamenáno, avšak bylo, např. pro mlhu, nepoužitelné, čemuž odpovídala symbolická hodnota používaná ve zprávě SYNOP pro tyto případy. Více než polovina takových případů pak připadá na noční termíny a na období s častými výskyty mlh. Velmi poznamenané jsou z tohoto ohledu referenční řady staničních pozorování typu oblačnosti. Situace je tím horší, čím se jedná o oblačnost vyššího patra, což je jasný důsledek zakrývání vyšší oblačnosti nižší a podrobněji bude tento jev rozebrán v podkapitole shrnující výsledky validací produktu CTY.

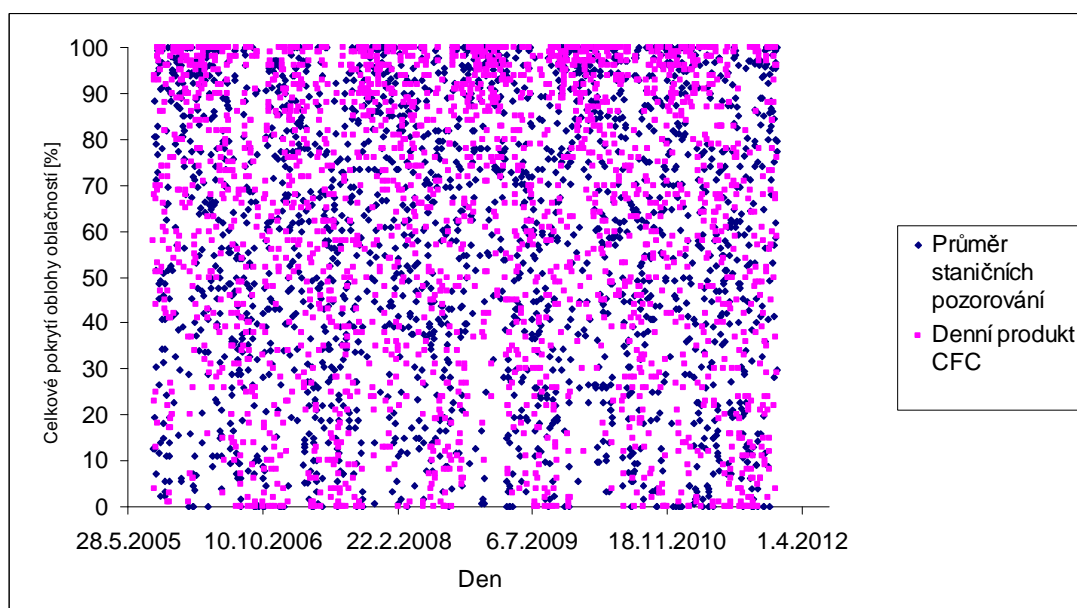
Kategorie typu oblačnosti	Korespondující kategorie CM-SAF	Korespondující kategorie SYNOP
Nízká oblačnost (dále ozn. jako <i>low</i> )	Opaque Low-level clouds	CL - Oblaky druhů St, Sc, Cu, Cb
Oblačnost středního patra (dále ozn. jako <i>middle</i> )	Opaque Medium-level clouds	CM - Oblaky druhů Ac, As, Ns
Oblačnost vyššího patra (dále ozn. jako <i>high</i> )	Opaque High-level clouds + Semi-transparent Cirrus clouds	CH - Oblaky druhů Ci, Cc, Cs + speciálně Cb

Tab. 2.3: Přirazení původních kategorií CTY a SYNOP k evaluovaným skupinám.

V předchozím odstavci popsané průměrování platí pro veličinu N, kterou je navíc ještě nutné převést z osmin na procenta. Pro veličiny ze zprávy SYNOP udávající typ oblačnosti je situace složitější, neboť se nejedná o vyjádření množství, ale zápis typu různých druhů oblaků. Možný validační koncept pak spočívá v převedení staničních pozorování ale i hodnot produktu CTY na binární vyjádření výskytu, nevýskytu daného typu. Následně je nutné vytvoření vlastních skupin typu oblačnosti, kam budeme rozdílné kategorie používané CM-SAFem a zprávou

SYNOP zařazovat. V této práci jsme se přidrželi rozdělení oblačnosti používaného Karlssonem (2008), viz tabulka 2.3, ovšem s tou změnou, že oblaky druhu Cb řadíme i do nízké i do vysoké oblačnosti. Dostaneme tak pro každý den tři binární údaje o výskytu oblaků kategorie low-level clouds (nízká oblačnost), medium-level clouds (oblačnost středního patra) a high-level clouds (vysoká oblačnost).

Máme tedy vytvořeny vzájemně si prostorově i časově odpovídající časové řady denních průměrů (resp. výskytů) vytvořené z referenčních staničních měření přidružené k časovým řadám MSG produktů. Vzhledem k určitým výpadkům, ať už ve staničních, tak i družicových řadách dat, je nutné pro každou oblast stanice v řadách denních průměrů a řadách denních výskytů typů odstranit ty dny, které chybí řadě s ní srovnávané. Příklad sobě si odpovídajících řad denních CFC a referenčních staničních denních průměrů je vizualizován v grafu 2.1. Za povšimnutí stojí malá shoda řad.



*Graf 2.1: Náhled na časové řady CFC a pokrytí oblohy oblačností v oblasti Kuchařovic.*

Zpráva SYNOP obsahuje výsledky přízemních meteorologických pozorování prováděných v celou hodinu. Oproti tomu MSG začíná snímkování zemského povrchu každých 15 minut, ale ani údaj s hlavičkovým časem měření v celou hodinu není pro území ČR přesně současný s pozemním měřením konaným v celou hodinu, neboť snímek našeho území je posunutý oproti hlavičkovému času zhruba o deset

minut. Porovnávané průměry tedy nejsou vytvořeny ze souborů se stejným množstvím prvků, ani se v nich nevyskytují žádná přesně současná měření. Možné důsledky této nedokonalosti našeho validačního souboru nastíníme v dalším odstavci.

Kvalita CM-SAF družicových produktů je pravidelně kontrolována a dle dokumentu CM-SAF (2011) jsou denní produkty z MSG dodávány s přesností na deset procent. Naproti tomu zkušený pozorovatel v denních podmínkách odhadne množství oblačnosti s přesností  $\pm 10\%$  a, i když je přesnost nočních pozorování omezena podstatně více než během dne, tak dle vzorců pro přenos chyby je chyba výsledného aritmetického průměru maximálně v řádech procent. Při porovnávání dvou odlišně průměrovaných řad se ale do celkové chyby promítá i chyba výběrová.

Oblačnost ve většině případů nesplňuje předpoklady stacionarity, což, vzhledem k již zmíněným obtížím správné prostorové kolokace družicových a staničních měření a vzhledem k nestejnému výběru měření do denních průměrů, zřejmě povede k tomu, že celková chyba mezi námi zkoumanými řadami hodnot denních CFC a CTY produktů a referenčních řad díky příspěvku výběrové chyby velmi naroste. Metodu pro odhad výběrové chyby pro produkt CFC představil Reuter a kol. (2010). Jejím použití v naší práci nicméně brání nedostupnost jednotlivých měření MSG, z kterých poté sestává výsledný denní produkt. V jedné z následujících podkapitol se tedy pokusíme alespoň minimalizovat výběrovou chybu tím, že do analýzy zahrneme pouze maximálně stacionární oblačné situace. To znamená, po uvážení dostatečné velikosti filtrovaného souboru, dny s denním průměrem pod dvacet a nad osmdesát procent pokrytí oblohy oblačností. Tyto dny budeme dále nazývat jako jasné či zamračené kvůli podobnosti s klimatologickou definicí těchto dnů.

## 2.2 Statistické metody

Námi vyšetřovaný problém má charakter analýzy časových řad. V podobě CFC a pokrytí oblohy oblačností pracujeme se spojitými omezenými veličinami, zatímco v podobě výskytu/nevýskytu kategorie typu oblačnosti či jasného nebo zamračeného dne se setkáváme s diskrétními veličinami s binárním rozdělením. Při analýze obou druhů časových řad budeme jako základní používat statistické metody doporučené WMO k validaci např. ve WMO TD No. 1485 (2009). U všech statistik bude uveden vzorec jejich výpočtu a stručně popsán význam.

## 2.2.1 Spojité veličiny

Do této kategorie spadají řady CFC a k nim referenční řady celkového pokrytí oblohy oblačností. Členy těchto řad jsou omezené, mohou nabývat pouze hodnot v uzavřeném intervalu nula až sto procent. Tyto řady byly analyzovány za použití následujících statistik při použití značení:

$P_i$ .....hodnota denního průměru veličiny odvozená ze staničních měření,

$D_i$ .....hodnota denního družicového produktu,

$X_i$ .....obecná veličina,

$N$ .....počet prvků řady.

$$\text{Průměrná hodnota (mean value): } MO = \frac{1}{N} \sum_{i=1}^N X_i . \quad (2.1)$$

Průměrná hodnota je užitečná pro lepší představu relativní velikosti odchylek a chyb a také je možno ji porovnat s klimatologickou hodnotou.

$$\text{Rozptyl (sample variance): } SVar = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2 . \quad (2.2)$$

Popisuje variabilitu veličiny.

$$\text{Směrodatná odchylka (sample standard deviation): } SSD = \sqrt{SVar} . \quad (2.3)$$

Popisuje variabilitu v jednotkách veličiny.

*Podmíněný medián (conditional median):* *Med*, poskytuje typickou hodnotu veličiny, přičemž je rezistentnější vůči odlehlým hodnotám než průměr.

*Mezikvartilová vzdálenost (interquartile range):* *ID*, je rovna vzdálenosti prvního a třetího kvartilu. V intervalu o této velikosti se vyskytuje 50% členů řady. Reflektuje variabilitu řady, přičemž je rezistentnější vůči odlehlým hodnotám než směrodatná odchylka.

$$\text{Systematická chyba (těž vychýlení, angl. bias): } ME = \frac{1}{N} \sum_{i=1}^N (D_i - P_i) . \quad (2.4)$$

Jedná se o veličinu reprezentující tendenci validovaného produktu systematicky nadhodnocovat ( $ME > 0$ ) či podhodnocovat ( $ME < 0$ ) pozorovanou veličinu oproti referenčním měřením.

$$\text{Střední absolutní odchylka (mean absolute error): } MAE = \frac{1}{N} \sum_{i=1}^N |D_i - P_i| . \quad (2.5)$$

Ukazuje střední hodnotu předpovědních chyb.



*Střední kvadratická chyba (mean squared error):*  $MSE = \frac{1}{N} \sum_{i=1}^N (D_i - P_i)^2$ . (2.6)

Značí průměr kvadrátu amplitudy chyby.

*Odmocnina ze střední kvadratické chyby (root mean squared error):*

$$RMSE = \sqrt{MSE}. \quad (2.7)$$

Jedna z nejčastěji používaných měř shody dvou souborů, díky obsaženému kvadrátu dává větší váhu vyšším hodnotám odchylek od skutečných hodnot, je též citlivější na výskyt odlehlých hodnot než *MAE*.

*Pearsonův korelační koeficient (correlation coefficient)* např:

$$r_p = \frac{\sum_{i=1}^N (D_i - \bar{D})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^N (D_i - \bar{D})^2} \sqrt{\sum_{i=1}^N (P_i - \bar{P})^2}}. \quad (2.8)$$

$r_p$  je také jednou z nejobvyklejších měř podobnosti dvou souborů hodnot. Velikost  $r_p$  je úměrná kovarianci  $D$  a  $P$ . Nabývá hodnot od mínus jedné do jedné, přičemž hodnotu jedné nabývá pro identické soubory. Je schopen postihovat pouze čistě lineární vztah mezi analyzovanými soubory, zatímco nelineární vazby jsou ignorovány a nejedná se o rezistentní veličinu, takže velikost  $r_p$  může být ovlivněna případnými odlehlými hodnotami.

## 2.2.2 Diskrétní veličiny

Vertikální rozsah a lokace oblačnosti se přímo podílí na rozdělení produktu CTY a staničních měření do výše v práci uvedených kategorií, dále označovaných jako *low*, *middle*, *high*. Pro družicový produkt i staniční měření máme tedy 3 binární řady denních výskytů kategorií typů. Schopnost správně detekovat tyto kategorie byla testována pomocí kritérií doporučených ve *WMO TD No. 1485 (2009)*. Nejprve byla data pro každou kategorii a stanici rozdělena do kontingenční tabulky, viz obr. 2.3, přičemž celkový počet případů  $n$  je roven:

$$n = a + b + c + d. \quad (2.9)$$

*Pozorované* znamená, zda se daná kategorie vyskytuje v referenční řadě, *Detekováno družicí* zastupuje výskyt jevu v řadách korespondujících družicových produktů.

		<i>Pozorované</i>		
		<i>Ano</i>	<i>Ne</i>	
<i>Detekováno : Ano družící</i>	<i>: Ano</i>	<i>a</i>	<i>b</i>	<i>a + b</i>
	<i>: Ne</i>	<i>c</i>	<i>d</i>	<i>c + d</i>
		<i>a + c</i>	<i>b + d</i>	

Obr. 2.3: Vzorová kontingenční tabulka.

Definice validačních kritérií jsou pak následující:

*Proportion correct:* 
$$PC = \frac{a+d}{n}. \quad (2.10)$$

Vyjadřuje podíl správně určených výskytů a nevýskytů oproti všem případům. Pokud je jedna z možností vzácná, může být toto kritérium lehce ovlivněno zvýšeným předpovídáním pravděpodobnější varianty. To by se ale mělo negativně projevit hlavně na *B* a *FAR* (viz dále).

*Threat score:* 
$$TS = CSI = \frac{a}{a+b+c}. \quad (2.11)$$

*TS* neboli také *Critical Success Index* je podílem správných detekcí oproti součtu všech detekcí a přehlédnutých výskytů. Je oproti *PC* vhodnější v případech, kdy se detekovaný jev vyskytuje zřídka.

*Vychýlení (bias):* 
$$B = \frac{a+b}{a+c}. \quad (2.12)$$

Frekvenční bias udává počet detekcí ku počtu skutečných výskytů. Může nabývat hodnot od nuly do nekonečna, ideální hodnota je rovna jedné.

*False alarm ratio:* 
$$FAR = \frac{b}{a+b}. \quad (2.13)$$

Jedná se o poměr falešných detekcí ku všem detekcím a tedy o kondicionální pravděpodobnost špatné detekce, pokud je jev detekován. Toto kritérium může být ovlivněno k lepšímu výsledku menším předpovídáním výskytu, což by se ovšem negativně odrazilo u *H*.

*Hit rate:* 
$$H = \frac{a}{a+c}. \quad (2.14)$$

Jedná se o pravděpodobnost detekce (*POD*), neboli kondicionální pravděpodobnost detekce za předpokladu výskytu jevu. Nabývá hodnot od nuly do

jedné, což znamená perfektní schopnost detekce. Dobrého výsledku v tomto kritériu může být dosaženo nadhodnocováním výskytu, což se ovšem projeví v biasu i *FAR*.

*False alarm rate:* 
$$F = \frac{b}{b+d}. \quad (2.15)$$

Jedná se o podíl nevýskytů, které byly detekovány jako falešný výskyt kategorie v daném dni. Je to tedy pravděpodobnost špatné detekce, podmíněná pravděpodobnost, že jev bude detekován i přesto, že se nevyskytuje.

Kombinací těchto kritérií se dále dá získat ještě *Heidkeho (HSS)* a *Pierceho skóre (PSS)*, která kvantifikují schopnost správného určení jevu družicí oproti náhodné předpovědi výskytu či nevýskytu zkoumané kategorie.

### 3 Analýza denního produktu CFC v oblasti ČR

Nyní přikročíme k prezentaci výsledků studia časových řad denního produktu CFC a jejich porovnání s referenčními řadami pokrytí oblohy oblačností ze stanic. Nejprve budeme zkoumat obecné vlastnosti řad, poté se budeme v jednotlivých podkapitolách pokoušet ověřovat hypotézy o možných příčinách odlišnosti od referenčních řad.

Řada CFC i referenční řada sice pokrývaly již zmíněné období od 1. září 2005 do konce roku 2011, avšak pro každou stanici jsou kvůli různým chybějícím dnům jinak dlouhé. Počet dnů v tomto období se spárovanými členy obou řad a tedy i počet analyzovaných členů každé z nich je uveden v tabulce 1.1. Můžeme si v ní všimnout viditelně menšího počtu analyzovaných dnů pro oblast Mošnova, což je zapříčiněno chybějícími dny v původní referenční řadě, kdy v listopadu nezřídka chybí údaje pro celé dny.

Stanice	Cheb	Kocelovice	Kuchařovice	Liberec	Mošnov	Pardubice	Přerov	Sedlec	Tušimice
N	2278	2270	2268	2276	2217	2277	2276	2264	2264

Tabulka 3.1: Délka řad pro jednotlivé oblasti.

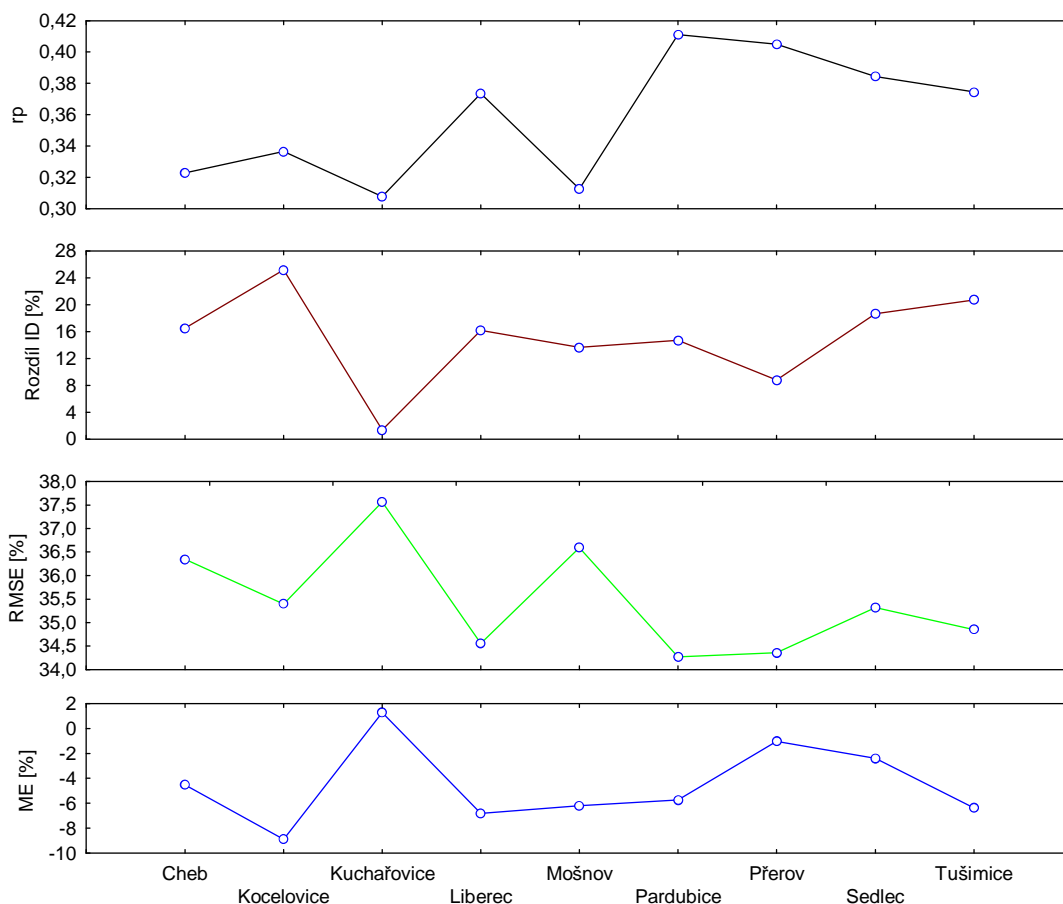
#### 3.1 Základní srovnání řady CFC a referenční řady

Základní charakteristiky časových řad denního produktu CFC a denních průměrů množství oblačnosti jsou uvedeny v tabulce 3.2. Tyto statistické ukazatele byly vypočítány pro každou oblast stanice zvlášť a poté i souhrnně pro řadu vytvořenou naskládáním členů řad z jednotlivých oblastí za sebou. Do tabulky 3.2 z důvodu přehlednosti nebyly zaneseny údaje ze všech stanic, ale jen z oblastí, které byly z hlediska shody mezi řadami ve většině ukazatelů nejextrémnější. Graf 3.1 naopak ukazuje, jak se mění ty nejzákladnější veličiny napříč zkoumanými oblastmi. Charakteristiky byly vybrány dle doporučení WMO TD No. 1485 (2009) a drží se zde uvedených definic.

Pro možnost porovnání celkové řady produktu CFC s celkovou referenční řadou, které by nám dalo souhrnné charakteristiky, je třeba udělat si představu, jak se řady liší oblast od oblasti. Zkoumání prostorové proměnlivosti shody řad bude ještě předmětem studia v následující podkapitole, nicméně již z grafu 3.1 si můžeme

udělat hrubou představu, jak ovlivníme výsledky, pokud řady složíme do jedné souhrnné řady pro produkt a jedné pro staniční průměry.

V níže uvedeném grafu si můžeme povšimnout, že *bias* i odmocnina ze *střední kvadratické chyby* se napříč stanicemi mění pouze v řádu procent. Zarážející je však velikost *RMSE* a znaménko *systematické chyby* v kontextu s výsledky práce Reutera a kol. (2009), k čemuž se blíže vrátíme za pár odstavců. Nyní chceme pouze odhadnout, jak reprezentativní budou charakteristiky celkového souboru. Bezrozměrná hodnota *Pearsonova korelačního koeficientu* se mění také nepatrně jen v rozsahu jedné desetiny, ale rozdíl *ID* je ovlivněn až příliš velkou podobností souborů pro oblast Kuchařovic, kde jsme celkově svědky odlišnějšího chování od ostatních oblastí ve všech čtyřech charakteristikách. I přesto by souhrnné hodnoty statistik měly mít dobrý smysl, jelikož většina stanic se vejde do rozmezí deseti procent, či jedné desetiny pro  $r_p$ .



Graf 3.1: Proměnlivost charakteristik shodnosti řad v závislosti na vyšetřované oblasti. Rozdíl ID značí  $ID_{cfc} - ID_{referenční}$ .

Hodnoty všech počítaných statistických veličin jsou uvedeny v tabulce 3.2. Kvůli přijatelným rozměrům tabulky jsme uvedli údaje jen pro stanice Pardubice, Kuchařovice a pro všechny stanice dohromady. Tyto dvě stanice nebyly vybrány náhodně, v Pardubicích je meteorologická stanice na letišti a měření zde jsou ve většině dnů prováděna i ve všech nočních termínech, zatímco v Kuchařovicích se měří pouze v omezené části dne. Z hodnot v tabulce 3.2 si můžeme udělat představu o tom, proč jsou charakteristiky shody řad v oblasti Kuchařovic oproti ostatním stanicím tak výjimečné (viz graf 3.1).

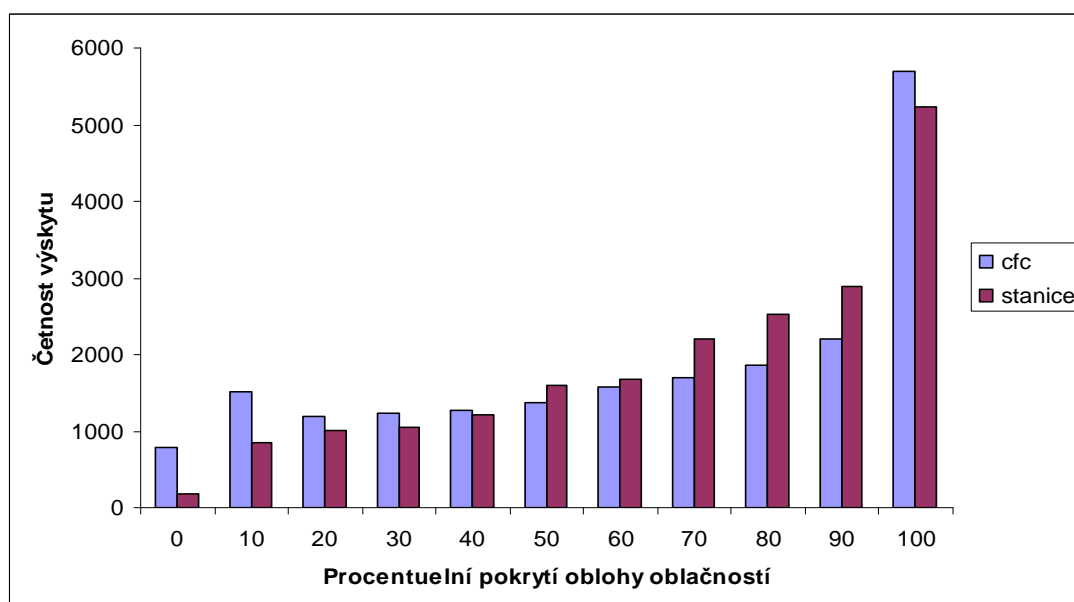
		<i>MO</i>	<i>SSD</i>	<i>SVAR</i>	<i>MED</i>	<i>ID</i>	<i>ME</i>	<i>MAE</i>	<i>MSE</i>	<i>RMSE</i>
<b>Pardubice</b>	stanice	66	29	829	72	45	-6	26	1175	34
	CFC	60	33	1097	67	60				
		<i>25thQ</i>	<i>75thQ</i>	<i>r<sub>p</sub></i>	<i>n</i>	<i>MO</i> -průměrná hodnota; <i>SSD</i> -směrodatná odchylka; <i>SVAR</i> -rozptyl; <i>MED</i> -podmíněný medián; <i>ID</i> -mezikvartilová vzdálenost				
	stanice	46	91	0,41	2277					
	CFC	32	92							
<b>Kuchařovice</b>		<i>MO</i>	<i>SSD</i>	<i>SVAR</i>	<i>MED</i>	<i>ID</i>	<i>ME</i>	<i>MAE</i>	<i>MSE</i>	<i>RMSE</i>
	stanice	61	31	992	65	55	1	29	1411	38
	CFC	62	32	1044	68	56				
		<i>25thQ</i>	<i>75thQ</i>	<i>r<sub>p</sub></i>	<i>n</i>	<i>ME</i> -systematická chyba (bias); <i>MAE</i> -střední absolutní odchylka; <i>MSE</i> - střední kvadratická chyba; <i>RMSE</i> - odmocnina střední kvadratické chyby				
	stanice	36	91	0,31	2268					
CFC	36	92								
<b>Souhrn pro všechny stanice</b>		<i>MO</i>	<i>SSD</i>	<i>SVAR</i>	<i>MED</i>	<i>ID</i>	<i>ME</i>	<i>MAE</i>	<i>MSE</i>	<i>RMSE</i>
	stanice	66	29	818	72	45	-5	27	1259	35
	CFC	61	33	1097	68	60				
		<i>25thQ</i>	<i>75thQ</i>	<i>r<sub>p</sub></i>	<i>n</i>	<i>25thQ</i> -hodnota prvního kvartilu; <i>75thQ</i> -hodnota třetího kvartilu; <i>r<sub>p</sub></i> – Pearsonův korelační koeficient				
	stanice	45	91	0,36	20400					
CFC	33	93								

Tab. 3.2: Základní charakteristiky řad. Hodnoty jsou uvedeny v procentech.

Jelikož v grafu 3.1 byly zaneseny pouze čtyři charakteristiky shody řad, nevyniklo to, že řada CFC má v této oblasti podobné statistické vlastnosti (*MO*, *SSD*, *SVAR*, *MED*, *25thQ*, *75thQ*) jako řady CFC v jiných oblastech. Referenční řada tu

má ale oproti jiným stanicím odlišné vlastnosti, poněvadž Kuchařovice leží dle Atlasu podnebí Česka (2007) v jedné z nejméně oblačných lokalit ČR. A protože je rozdělení hodnot referenční řady v této oblasti podobnější obecnému rozdělení CFC, ne tedy kvůli specifické vlastnosti řady CFC v oblasti Kuchařovic, můžeme tak pochopit nestandardně malý *rozdíl ID* vykazovaný v této oblasti v hořejším grafu. Prostorové variabilitě se budeme více věnovat v samostatné podkapitole.

Veličiny *Med*, *ID* a první a třetí kvartil naznačují, že v řadě *cfc* je více dnů s malým pokrytím oblačností a více dnů s pokrytím oblačností větším než 90%, což dobře dokumentuje graf 3.2. Zajímavé je, že rozdělení CFC se mnohem méně podobá teoretickému gama rozdělení typickému pro omezené jevy.



Graf 3.2: Četnosti výskytů hodnot denních průměrů.

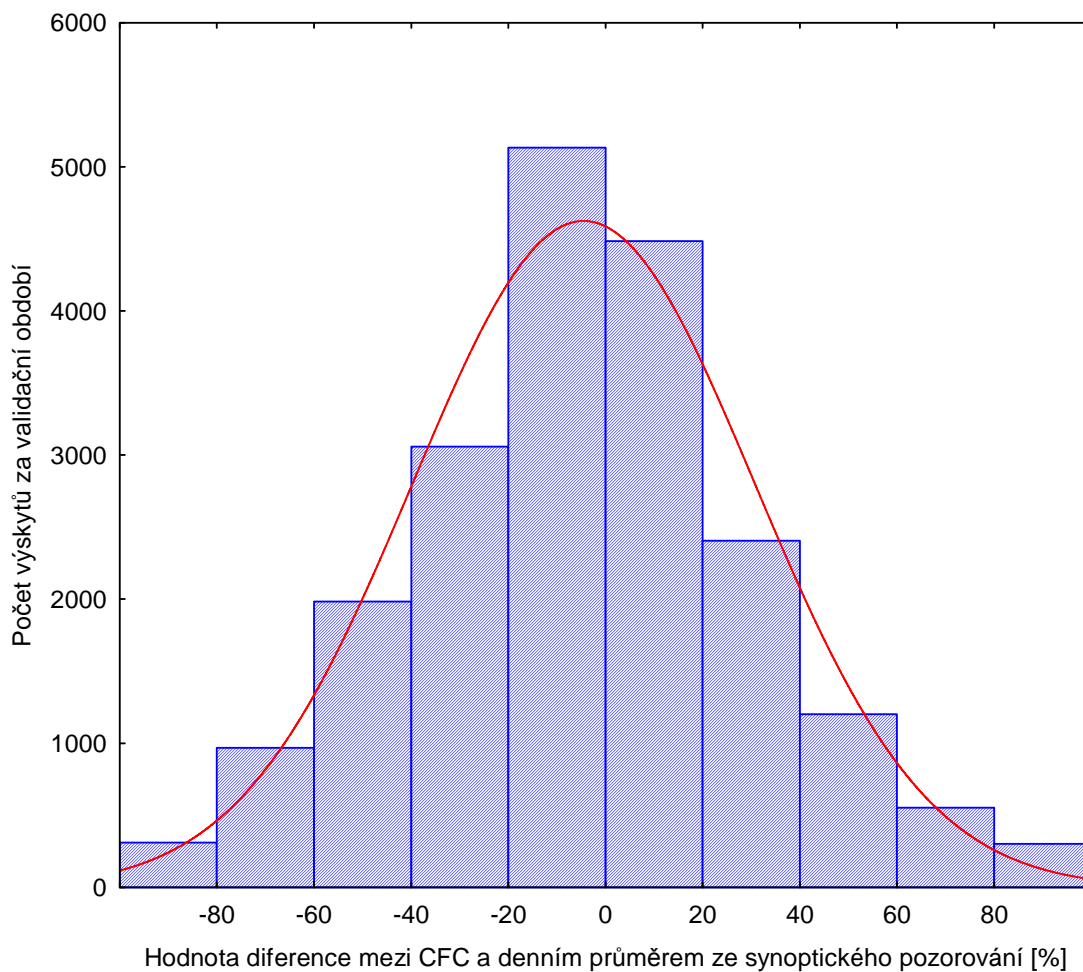
Velikost biasu vypovídá o mírném systematickém podhodnocování denního průměru množství oblačnosti produktem CFC. Shoda produktové s referenční řadou je však mnohem horší, jak dokládají vysoké hodnoty veličiny MAE případně RMSE. Reuter a kol. (2009) validovali ve své práci mimo jiné také denní průměry CFC. V jejich studii byly ale nejprve vypočítány vlastní hodnoty denního CFC tak, že z měření MSG brali pro celý MSG disk pouze hodnoty prostorově (pracovali přímo v pixelovém rozlišení) a časově nejbližší stanici a pozorování na ní, takže nevyužili všechna měření MSG. Tím zmenšili výběrovou chybu. Pro pevninu došli k biasu 5% nadhodnocování a jejich *RMSE* je rovna 17%. V této práci se ovšem jedná o

odmocninu střední kvadratické chyby korigované o bias. Tato korekce pro naše data nevedla k jednoznačnému zmenšení *RMSE*, proto od ní bylo upuštěno. V tomto světle vypadá námi vypočítaný záporný bias podivně (nutno podotknout, že při validaci CFC výsledky ze senzoru MODIS dospěli Reutera a kol. (2009) také k záporné systematické chybě). Podhodnocování a větší hodnota *RMSE* v naší studii lze pravděpodobně přičíst výběrové chybě. Speciálně podhodnocování bude zřejmě způsobeno větším množstvím nočních měření promítajících se do řady CFC než do řady referenční, protože dle standardního denního chodu množství oblačnosti připadá na noc minimum oblačnosti. Pokud by tomu tak bylo, měli bychom v další podkapitole zabývající se prostorovou proměnlivostí řad a jejich shody najít mezi oblastmi rozdíly ve velikosti *ME*. V oblastech stanic s nekompletními nočními pozorováními by měla být systematická chyba větší.

Hodnota Pearsonova korelačního koeficientu je od jedné, tedy absolutní shody, dosti vzdálena, což svědčí v konkrétních dnech o nepřilíš velké podobnosti hodnot členů řad CFC a denního průměru pokrytí oblohy oblačností.

Reuter a kol. (2009) také ukázali, že difference CFC od denních průměrů pokrytí oblohy oblačností ze synoptických pozorování mají přibližně normální rozdělení s posunem do kladných hodnot kvůli nadhodnocování od CFC. Pro naše data ale nebylo dosaženo podobného výsledku, neboť množství výskytů neubývá směrem k okrajovým hodnotám tolik, jak by příslušelo normálnímu rozdělení, viz graf 3.3, kde je patrný také vliv podhodnocování převažující ve všech intervalech hodnot kromě těch s v absolutní hodnotě maximální diferencí.





Graf 3.3: Histogram difference proložený křivkou normálního rozdělení.

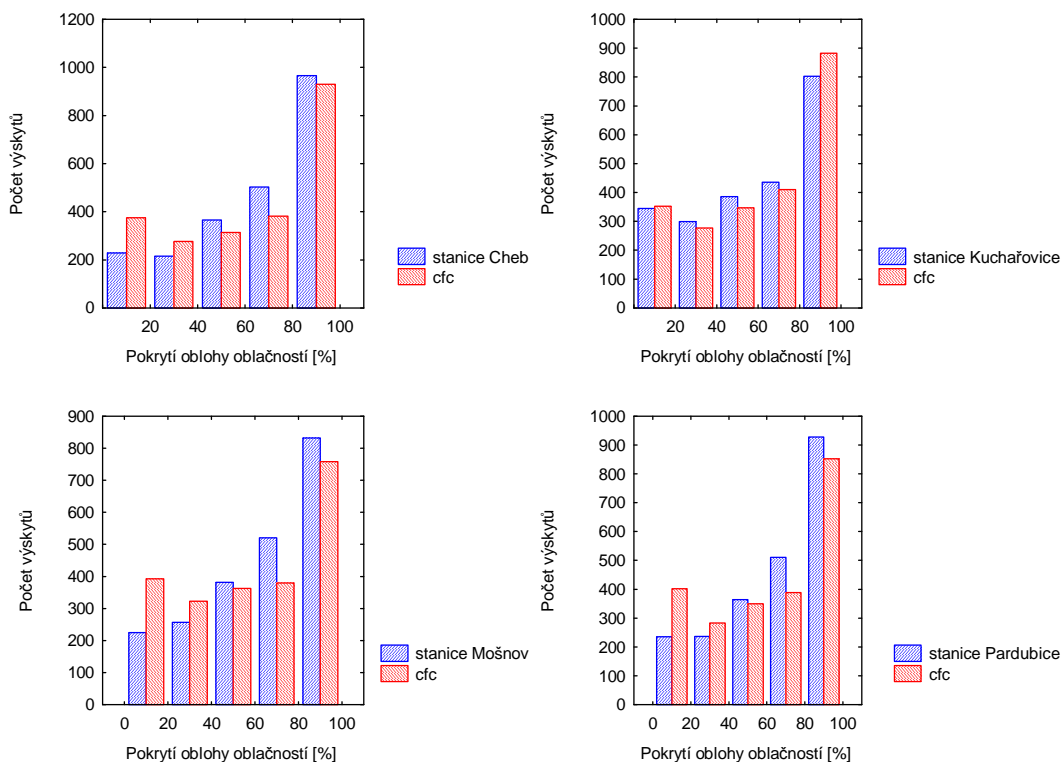
### 3.2 Prostorová variabilita shody řady CFC a referenční řady

Nejdříve je třeba si blíže všimnout oblastí, v kterých řady vyšetřujeme. Všechny stanice mají nadmořskou výšku mezi 200 a 500 m n.m. (viz tabulka 2.2). Žádná ze stanic není horská, kde by se dal očekávat jiný roční chod oblačnosti a obtíže produktu v závislosti na použité zjednodušené orografii. Pouze v Liberci se vyskytuje okolí s výraznější topografií v podobě Ještědu a Jizerských hor, pravděpodobně ale ne dost výrazné, aby se v naší analýze projevilo. Z obrázku 2.1 je patrné rozmístění zkoumaných oblastí prakticky po celé ploše České republiky. Dle Atlasu podnebí Česka (2007) by klimatologicky všechny lokality měly mít podobné průměrné množství oblačnosti, jen jihomoravské stanice Sedlec a Kuchařovice spadají do oblasti s menší oblačností.

V této podkapitole budeme také studovat, jaký vliv na míru podobnosti řady CFC s referenční má režim měření na stanici. Jestli tedy budou v okolí stanic Kuchařovice, Tušimice a částečně také v oblastech Liberce a Kocelovic výsledky odlišné (horší) od okolních stanic. Částečnou odpověď nám nabízí již graf 3.1 z předchozí podkapitoly, zaměříme-li se v něm na veličiny  $RMSE$  a  $r_p$ , které přímo charakterizují odlišnost řad člen po členu, zjistíme, že nejhorší shoda dle obou charakteristik panuje opravdu v oblasti Kuchařovic, ale naopak Tušimice a Liberec dosahují lepších výsledků než většina ostatních lokalit. Nejlépe odpovídá řada CFC referenční v oblasti Pardubic a Přerova, kde velikost korelačního koeficientu přesahuje hodnotu 0,4. Může to být tím, že pozorování na těchto stanicích patří k nejhomogennějším z celého souboru s pravidelným plným zastoupením nočních pozorování. Tuto domněnku však do jisté míry vyvrací Mošnov se stejným množstvím nočních měření, kde, vedle oblasti Kuchařovic, si řady odpovídají nejméně.

V grafu 3.4, do kterého jsou zahrnuty jak oblasti s nejhoršími podobnostními charakteristikami z grafu 3.1, tak pro porovnání s nimi i Pardubice, lze vidět, že rozdělení hodnot produktu je velice podobné pro všechny oblasti. Při interpretaci grafu je třeba si uvědomit, že v každé oblasti mají řada CFC a řada referenční jiné množství členů. Je vidět, že co se týče rozdělení hodnot, tak produkt CFC lépe vystihuje region s menší oblačností jako Kuchařovice, jinde je nadhodnoceno množství jasných dnů i dnů z navazujících kategorií s průměrným pokrytím oblohy oblačností až do 40%. Počet zamračených dnů je naopak všude kromě lokality Kuchařovic podhodnocován.

Příčinu můžeme hledat ve výběrové chybě, i když je zarážející, že se stejně tak děje i v Pardubicích s pravidelnými nočními měřeními, kde noční minimum přispívá k celkovému dennímu průměru stejnou vahou jako u produktu CFC. Budeme tento jev studovat ještě v další podkapitole, kde se pokusíme najít vliv ročního chodu, což by mohlo vést k osvětlení možných důvodů. Prozatím zůstává otázkou, proč je rozdělení hodnot produktu CFC natolik odlišné od rozdělení hodnot pokrytí oblohy oblačností převládajícího na většině stanic.



Graf 3.4: Četnosti výskytů hodnot řad ze čtyř oblastí.

Abychom se pokusili zmenšit vliv výběrové chyby v podobě mnohem častějšího měření MSG, budeme nyní analyzovat dny s množstvím oblačnosti pod dvacet a nad osmdesát procent, pro podobnost s klimatologickou definicí označené jako jasné a zamračené dny. V těchto dnech je největší pravděpodobnost přibližné stacionarity oblačnosti. Budeme zkoumat schopnost CFC produktu „předpovědět“ takovéto dny, tedy udávat je za předpokladu, že se vyskytují dle staničního průměru, a budeme se samozřejmě zajímat o prostorovou variabilitu této schopnosti. Jelikož se jedná o 2 binární problémy výskytu, nevýskytu jasného a zamračeného dne, využijeme k analýze validačních kritérií představených v sekci 2.2.2

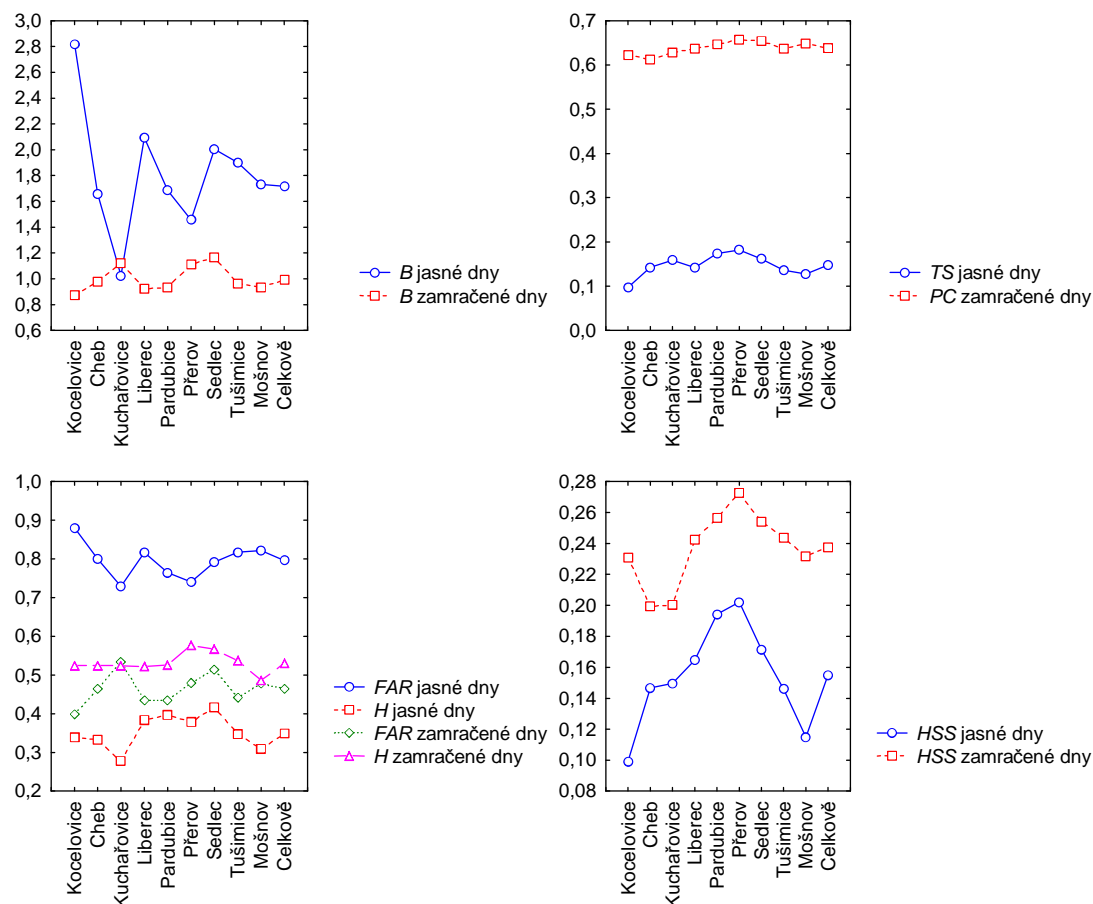
Údaje z tabulky 3.3 uvedené níže můžeme ještě před tvorbou kontingenční tabulky a výpočtem validačních kritérií porovnat s rozložením jasných a oblačných dnů na území ČR dle Atlasu podnebí Česka (2007). Metoda výpočtu těchto dní je v Atlasu sice jiná, nás ale nebude tak zajímat číselné množství, jako spíše prostorové rozložení počtu dnů na našem území, což již půjde porovnat s údaji v tabulce 3.3. Zatímco podle Atlasu podnebí Česka (2007) ubývá na našem území počet jasných dnů s rostoucí zeměpisnou šířkou a zmenšující se délkou, tak produkt CFC neumí tuto závislost přesně zachytit. Při interpretaci je třeba dát si pozor na různý počet

dnů, z kterého jsou jasné a zamračené dny počítány v jednotlivých oblastech. Nicméně je zřetelné příliš malé množství jasných dnů uváděné produktem CFC pro oblast Kuchařovic v kontextu geografické polohy a počtu jasných dnů v řadách CFC v jiných oblastech. V blízkém Sedlci už údaj vypadá věrohodněji. Obecně se počet jasných dnů produktu CFC příliš nemění, prostorová variabilita je rozhodně menší oproti jasným dnům počítaným z referenční řady.

Oblast	Počet analyz. dnů	Stanice jasný den	Stanice: zamračený den	CFC jasný den	CFC zamračený den	Počet shod jasný den	Počet shod zamračený den
Kocelovice	2270	133	1043	375	910	45	547
Cheb	2278	220	950	365	930	73	498
Kuchařovice	2268	328	785	336	882	91	411
Liberec	2276	169	939	354	867	65	490
Pardubice	2277	229	916	386	853	91	482
Přerov	2276	254	819	370	908	96	473
Sedlec	2264	218	760	437	886	91	431
Tušimice	2274	190	932	361	897	66	501
Mošnov	2217	220	814	381	758	68	396
<i>Celkově:</i>	<i>20400</i>	<i>1961</i>	<i>7958</i>	<i>3365</i>	<i>7891</i>	<i>686</i>	<i>4229</i>

*Tab. 3.3: Počty jasných a zamračených dnů dle oblastí.*

U zamračených dnů je podle Atlasu podnebí Česka (2007) situace s prostorovým rozložením spíše opačná, více se zde projevuje vliv pohoří na růst počtu těchto dnů. Na první pohled nás v tabulce 3.3 zaujme i přes menší počet vyšetřovaných dnů příliš malé množství zamračených dnů dle CFC produktu pro oblast Mošnova. Celkově je v produktu CFC prostorová variabilita počtu zamračených dní velmi malá a nevystihuje hlavní klimatologické znaky jako minimum na jihu Moravy. V referenční řadě tyto znaky vidět jsou.



Graf 3.5: Souhrn výsledků vybraných validačních kritérií řad ze všech oblastí.

V grafu 3.5 můžeme vidět výsledky vybraných validačních kritérií, které by nám měly pomoci zhodnotit schopnost řad produktu CFC vystihnout v daných oblastech chování referenčních řad. Budeme předpokládat, že oblačnost v jasných a zamračených dnech bude do velké míry stacionární a v těchto dnech bude tedy časová i prostorová složka výběrové chyby hrát menší roli.

Z části grafu obsahující hodnoty Heidkeho skóre si ihned můžeme udělat představu, že schopnost správné předpovědi těchto dnů produktem CFC opět není moc dobrá, přičemž u zamračených dnů je přece jen zhruba o desetinu lepší. Podíváme-li se, jak se tato schopnost liší oblast od oblasti, zjistíme, že nejlepší výsledky mají stejně jako v sekci 3.1 oblasti Pardubic a Přerova. Kuchařovice s průměrnou hodnotou *HSS* dokládají, že schopnost správné detekce v určitý den není moc ovlivněna dobrým vystižením celkové frekvence výskytu dané hodnoty, kdy (viz graf 3.4 a tabulka 3.3) je v této lokalitě nejmenší rozdíl mezi celkovým pozorovaným a detekovaným množstvím jasných dnů.

Tuto schopnost udávání shodného celkového počtu dnů kvantifikuje veličina  $B$  a graf 3.5 jen dokumentuje naše předešlé tvrzení, že produkt CFC ve všech lokalitách vyjma Kuchařovic znatelně nadhodnocuje počet jasných dnů. Nejextrémněji, až trojnásobně v oblasti Kocelovic. Počet zamračených dnů je oproti referenční řadě lehce nadhodnocený pro moravské oblasti, ve zbytku lokalit spíše podhodnocený. Hodnoty vychýlení jsou ale na celém území téměř zanedbatelné, blízké jedné.

Podíl správných detekcí výskytu či nevýskytu je pro jasné dny, jakožto pro řídkce se vyskytující jev, vyjádřen v grafu 3.5 veličinou  $TS$ , zatímco pro mnohem běžnější zamračené dny veličinou  $PC$ . Co se prostorové variability týče je zde u obou typů dne vidět nevýrazné maximum úspěšnosti v lokalitách Pardubic a Přerova. Celkově lze vidět, že ve výskytu zamračených dnů je mezi oběma řadami mnohem větší shoda než u jasných dnů, což ale mnohem přesněji hodnotí již diskutované Heidkeho skóre.

Poslední část grafu 3.5 ukazuje do jisté míry spojené veličiny  $H$  a  $FAR$ . Prvně jmenovaná počítá s výskytem jevu v referenční řadě a udává pravděpodobnost, že se toho dne jev vyskytuje i v řadě CFC.  $FAR$  značí pravděpodobnost, že jev, pokud se vyskytuje v řadě CFC, nebude v referenční řadě. Čím je tedy velikost  $FAR$  menší, tím lepší je schopnost předpovědi jevu. Pro jasné dny ovšem vidíme vysoké hodnoty  $FAR$  a tedy velké procento falešných detekcí, nejvíce v oblasti Kocelovic. Nejvíce falešných detekcí zamračených dnů a to každý druhý výskyt máme pro jihomoravské stanice. Nejmenší *hit rate* má pro oba druhy dnů Mošnov a pro jasné dny Kuchařovice.

Celkově lze říci, že schopnost produktu CFC ve zkoumaných stacionárních situacích dávat podobné hodnoty (strefit se udávaným množstvím oblačnosti do krajních intervalů velikosti 20%) jako referenční řada je ze zatím neznámých příčin malá, lepší pro zamračené dny (denní průměr pokrytí oblohy oblačností 80% a více). V jasných dnech, tedy dnech téměř bez oblačnosti, je tato schopnost produktu CFC v některých lokalitách jen o trochu lepší než náhodná předpověď výskytu/nevýskytu jasného dne. Zatímco možné příčiny těchto neshod budeme ještě dále hledat, nyní se pokusíme odpovědět na otázku prostorové variability produktu CFC a shody s referenční řadou.

Zjistili jsme, že CFC prostorovou variabilitu množství oblačnosti na území ČR sice reflektuje, mnohdy však trochu nepřesně a hlavně co do velikosti proměnlivosti značně utlumeně. Řady produktu CFC mají ve všech oblastech velmi podobný

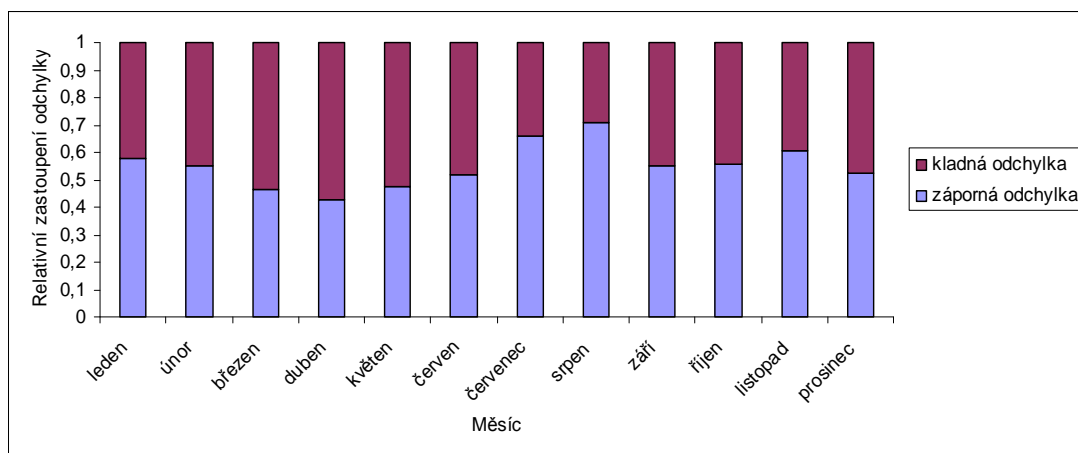
charakter. Nejlepší shody je dosaženo většinou u stanic s nejpravidelnějšími měřeními a bez význačné geografické polohy jako např. Pardubice a Přerov. Nejhůře se produkt CFC shoduje s průměrem staničních pozorování v případě jasných dnů, jejichž počet významně nadhodnocuje hlavně pro české stanice.

### 3.3 Časová závislost kvality produktu CFC

Jeden z dílčích cílů, který jsme si na počátku práce určili, bylo ověřit, zda se kvalita denního produktu CFC mění významně s časem např. ve spojitosti s ročním chodem oblačnosti. Pro chybějící dny ve vyšetřovaných časových řadách a tedy nespojitému ročnímu chodu nemohly pokročilejší metody typu Fourierovy a Wavelet transformace dávat směřodátne výsledky. Proto jsme řady ve všech oblastech analyzovali chronologicky po celé období v měsíčních intervalech. Pro každý měsíc od září roku 2005 do konce roku 2011 byl vypočten měsíční průměr produktu CFC i denních průměrů staničních pozorování a byly získány charakteristiky vlastností a podobnosti těchto řad v daném měsíci v každé lokalitě. To samé bylo uděláno i pro řadu CFC a referenční řadu souhrnné přes všechny oblasti.

Předpokládáme, že díky tomu budeme schopni lépe identifikovat příčiny odlišností, na které jsme narazili v předchozích podkapitolách. Například v případě horších výsledků zimních měsíců bychom mohli větší odlišnost řady CFC od referenční řady přisuzovat většímu podílu pozorování po setmění v denním průměru v tomto období, nebo výskytu pro družice hůře detekovatelných scén nad zamrzlým terénem či mlh. V případě horších letních či jarních výsledků budeme zase moci lépe zhodnotit vliv nestacionarity oblačnosti, kdy v tomto období častá konvektivní oblačnost může způsobit rozdílné hodnoty mezi jedním pozorováním za hodinu ze stanice a čtyřmi detekcemi za hodinu od MSG, atd.

V části 3.1 jsme zjistili, že produkt CFC oproti denním průměrům ze staničních pozorování systematicky podhodnocuje pokrytí oblohy oblačností. To však ještě automaticky neznamená, že ve většině dnů by produkt CFC měl udávat menší oblačnost než referenční průměr. Vývoj počtu dní s kladnou (CFC udává vyšší hodnotu než průměr staničních pozorování) či zápornou odchylkou (menší hodnota CFC) v závislosti na měsíci je znázorněn v grafu 3.6.



Graf 3.6: Relativní zastoupení kladných a záporných odchylek souhrnné řady CFC od řady referenční.

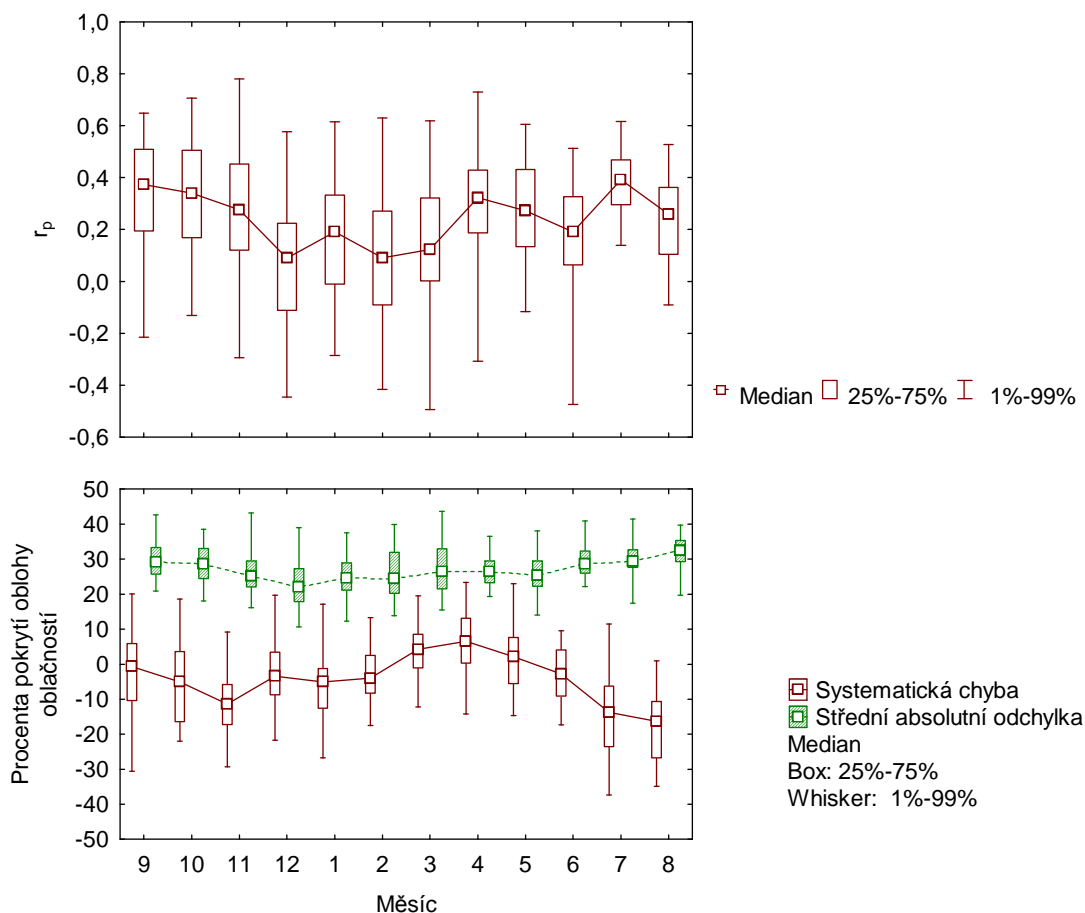
Můžeme si povšimnout, že opravdu ve většině srovnávaných dnů ze všech oblastí celkově udával denní produkt CFC menší procento pokrytí oblohy oblačností než referenční denní průměr. Největší podíl těchto dnů nebyl překvapivě v zimních měsících s větší vahou noční části denního chodu oblačnosti, ale v červnu a v srpnu. V druhém jmenovaném měsíci dokonce produkt podceňuje oblačnost ve více jak sedmdesáti procentech dnů. Naopak jedině v jarních měsících je nepatrně více dnů s kladnou odchylkou než se zápornou. Mohlo by to být vysvětleno tím, že je v těchto měsících oblačnost velmi proměnlivá, znatelně nestacionární a přitom velmi často konvektivního charakteru. Na rozdíl od léta se však často jedná o oblaky spíše malých až podpixelových rozměrů, jejichž výskyt prokazatelně vede k nadhodnoceným denním průměrům, viz podkapitola o CFC produktu.

Velikost systematické chyby v závislosti na měsíci můžeme najít v grafu 3.7 spolu s vývojem hodnoty Pearsonova korelačního koeficientu a střední absolutní odchylky, která byla vybrána jako měřítko střední velikosti chyby místo *RMSE* pro zmenšení vlivu případů s velkými odchylkami. S tím koresponduje i volba prezentace výsledků pomocí krabicového diagramu s vousy a s hodnotou mediánu místo průměru. Jednotlivé hodnoty spadající do jednoho měsíce jsou totiž (vzhledem k původu z různých lokalit a z různých let) od sebe často velmi vzdáleny.

Hodnoty biasu v grafu 3.7 ve svém vývoji kopírují roční chod zachycený v grafu 3.6, ovšem v krabicovém diagramu se můžeme přesvědčit o velké proměnlivosti systematické chyby jednotlivých měsíců napříč celým souborem. Zejména v červenci, srpnu, září a říjnu přesahuje pro systematickou chybu velikost boxu (z níž přímo vyplývá veličina *ID* popsaná v části 2.2.1) 15% pokrytí oblohy



oblačností. Ve většině měsíců navíc vousy ohraničující výskyt velikostí biasu s pravděpodobností alespoň jednoho procenta zasahují hluboko do kladných i záporných hodnot, takže lze vidět jeho velkou proměnlivost. Nejkoncentrovanější hodnoty systematické chyby má červen. U něj stejně jako u několika dalších měsíců nelze vyloučit vymizení systematické chyby při středování přes delší než šestiletý soubor.



Graf 3.7: Krabicové diagramy ročního chodu systematické chyby, střední absolutní odchylky a Pearsonova korelačního koeficientu.

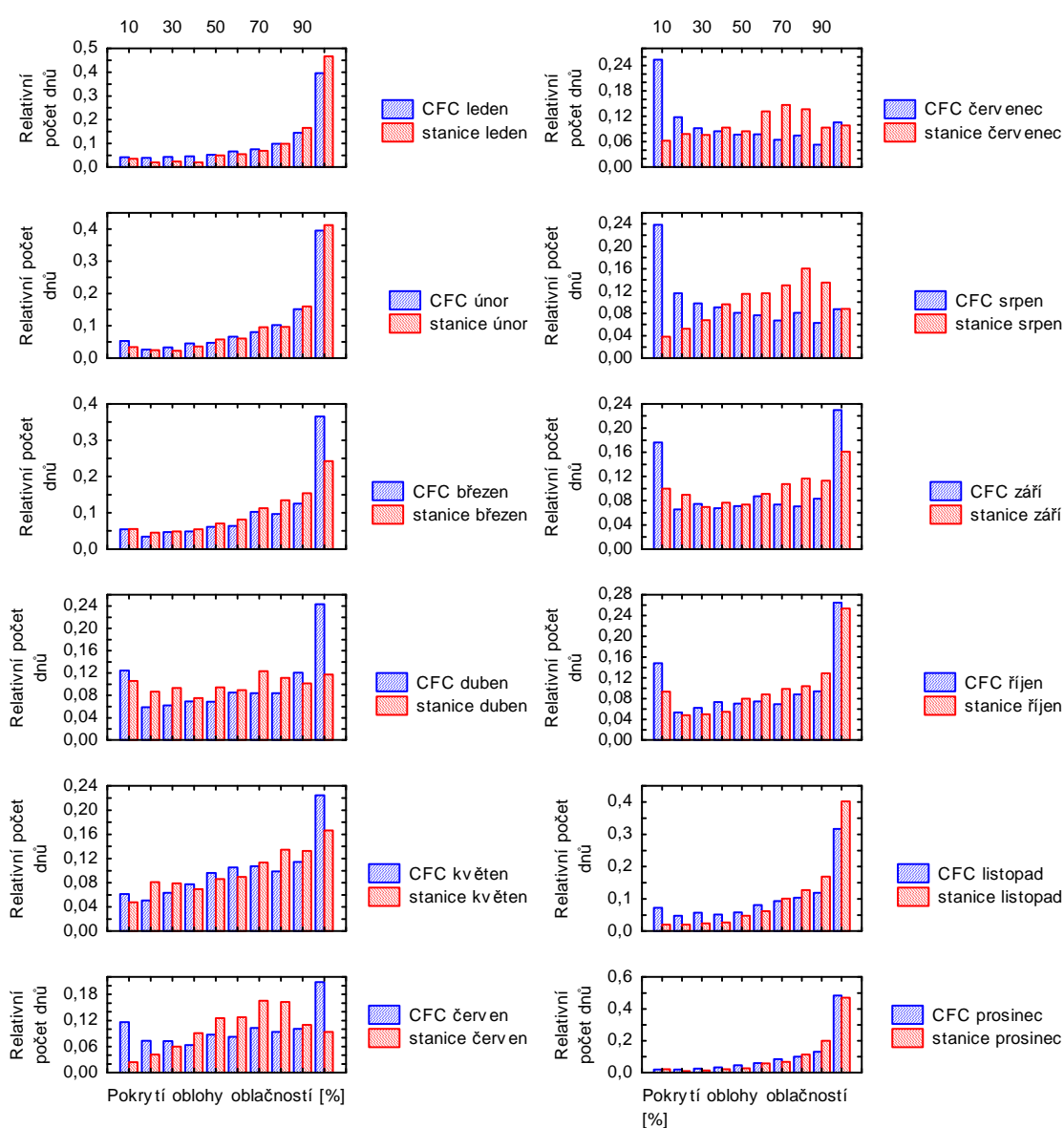
Roční chod střední chyby denního produktu CFC je, jak vidíme v grafu 3.7, téměř nezatelný. Hodnota střední absolutní odchylky všech měsíců se v 98% případů nachází mezi 20 a 40% procenty pokrytí oblohy oblačností. Znatelný i když velmi nepravidelný roční chod vykazuje medián hodnoty Pearsonova korelačního koeficientu v horní části krabicového diagramu.

Nejhorší hodnoty shody mezi řadami jsou v prosinci, únoru a březnu, jen o málo lepší je shoda v lednu a překvapivě v červnu, kde je podobnost řad absolutně nejhorší z celé letní části roku. Měsíce v zimní polovině mají celkově nižší hodnoty korelace a i rozptýlenost těchto hodnot je větší. Boxy  $r_p$  v prosinci, únoru a částečně i lednu zasahují i do záporných hodnot korelace, což by značilo, že v některých případech nabýval vztah ( $r_p$  je schopno postihnout jen lineární závislosti) mezi řadou družicového produktu a staničních průměrů v celém měsíci, ač nepatrně, charakteru nepřímé závislosti. Bude to zřejmě důsledkem opačné klasifikace některých situací družicovým produktem a pozorovatelem. Například při tenké oblačnosti tvořící se při hranici výškové inverze, kterou by družice nezaznamenala a pozorovatel ano. Tento příklad je však pouze demonstrační a skutečná příčina negativní korelace může být jiná. Obecně horší shodu řad v zimních měsících vysvětlujeme při přihlédnutí k sekci 1.4.1 obtížnější detekcí oblačnosti v těchto měsících, ať už kvůli vyššímu výskytu obtížně detekovatelných mlh, nízké a inverzní oblačnosti, či kvůli slabému teplotnímu kontrastu studeného zemského povrchu vůči oblačnosti. Také se zde zřejmě projeví dříve zmíněný vliv většího podílu nočních pozorování, která jsou zatížena větší chybou detekce na jedné a pozorování na druhé straně, což spolu s větším počtem těchto měření zahrnutým v souhrnné řadě CFC se posléze projeví i v denním průměru a větší výběrové chybě.

Největší korelace mezi řadami je v červenci. Bráno podle nejvyšší hodnoty mediánu, nebo hodnoceno dle nejvyšší koncentrovanosti vyšších hodnot koeficientu  $r_p$ , jelikož v červenci mají vousy i boxy nejmenší délku a jsou přitom položeny v nejvyšších hodnotách. Důvod odlišného chování července oproti červnu a srpnu nejsme prozatím schopni vysvětlit.

Roční chod charakteristik shody obou řad je velmi nepravidelný a navíc nejsme schopni uspokojivě objasnit příčiny pozorovaných rysů. Pomocné informace bychom mohli získat z porovnání ročního chodu nabývaných hodnot obou řad, proto je uveden graf 3.8, kde je možné pro každý měsíc pozorovat a porovnávat rozdělení velikosti CFC a hodnoty řady denních průměrů pokrytí oblohy oblačností (opět souhrnně po celé období a pro všechny oblasti dohromady). Jedná se o histogram relativních zastoupení jednotlivých kategorií hodnot pokrytí oblohy oblačností v daném měsíci, protože každý měsíc má (vlivem různých délek, chybějících dat a kvůli odlišnému počtu zastoupení v našem období) různý celkový počet dnů.

Pokud si v grafu 3.8 projdeme postupně měsíc po měsíci rozdělení hodnot průměrů ze staničních pozorování, směrem po časové ose od ledna vidíme postupné přelévání počtu výskytů od vyšších hodnot pokrytí oblohy oblačností k nižším, téměř rovnoměrné rozdělení v dubnu, opět rostoucí zastoupení dnů s větší oblačností v květnu, načež přicházejí letní měsíce s maximem zastoupení v středních hodnotách, které se v jiných měsících nevyskytuje. Září má spolu s dubnem a říjnem největší zastoupení jasných dnů, poté už však rozdělení začíná znovu pomalu nabývat tvaru typického pro zimní měsíce.



Graf 3.8: Histogram znázorňující rozdělení hodnot souhrnné řady CFC a řady referenční.

Rozdělení hodnot produktu CFC právě v zimních měsících dobře vystihuje rozdělení referenční řady. V lednu a únoru CFC udává o trochu méně zamračených dnů, v prosinci nepatrně více dnů s pokrytím oblohy oblačností vyšším než devadesát procent. V druhé nejvyšší procentuální kategorii je však rozdíl opačný. V březnu, dubnu a květnu se v řadě produktu CFC vyskytuje mnohem více dnů s oblačností devadesát procent a vyšší. Děje se tak částečně na úkor dnů ze středních kategorií hodnot, nicméně v nadbytku dnů tohoto maximálního množství oblačnosti můžeme spatřovat příčinu nepatrně kladného biasu v grafu 3.7. Jinak se rozdělení hodnot CFC v základních rysech přibližně shoduje s referenčním, správně byl detekován menší počet dnů s minimem oblačnosti v dubnu než v květnu.

V letních měsících vidíme v grafu 3.8 naprostý nesoulad mezi rozdělením hodnot CFC a referenčním. Zejména v červenci je to o to překvapivější, že výše diskutované charakteristiky shody zde byly ze všech měsíců nejlepší a potvrzuje to naši domněnku, že v naší úloze se shodnost rozdělení hodnot a shodnost jednotlivých členů řad příliš neovlivňují.

V červnu, červenci a srpnu vidíme v řadě CFC značný nadvýskyt jasných dnů, a naopak méně dnů s hodnotou pokrytí oblohy oblačností 50-90%. V červnu je pak v řadě CFC dvakrát více dnů s maximální oblačností než v referenční řadě. V červenci a srpnu si jsou rozdělení hodnot produktu navzájem velmi podobná s větším množstvím dnů v malých hodnotách pokrytí oblohy oblačností, což vysvětluje zápornou systematickou chybu pozorovanou v grafu 3.7.

Fyzikální příčina této chyby není ovšem zcela jasná. Podle sekce 1.4.1 bychom očekávali od produktu v letních měsících spíše nadhodnocování spojené s výskytem konvektivní podměřítkové oblačnosti. Možnou příčinou by mohl být větší podíl nočních měření v souhrnné řadě denního CFC oproti referenční řadě, jelikož v těchto měsících je pokrytí oblohy oblačností v nočních hodinách výrazně menší než ve dne. Vysvětlení, ke kterému bychom se klonili nejvíce, však požaduje nabourání použitého validačního konceptu, kdy bereme staniční pozorování za vyjádření pravdivého stavu atmosféry. Pozorovateli se totiž konvektivní oblačnost vlivem geometrie problému s rostoucí vzdáleností od něj spojuje, což může vést k nezanedbatelné kladné systematické chybě zahrnuté v staničním měření. Tento jev by pak mohl přispět k menším hodnotám produktu CFC oproti referenční řadě v červenci a v srpnu a k vysvětlení, proč v jarních, letních a částečně i podzimních měsících má řada CFC oproti referenční více dnů s minimální oblačností.

V září a říjnu mají rozdělení hodnot CFC oproti referenčním rozdělením větší zastoupení v obou krajních intervalech. V listopadu je v řadě CFC více jasných a méně zamračených dnů. V prosinci si jsou obě rozdělení podobná a v rámci řady CFC je prosinec měsícem s největším výskytem denního průměru pokrytí oblohy oblačností větším než 90%.

### **3.4 Závislost kvality vybraných družicových produktů na typu oblačnosti**

V této podkapitole budeme zkoumat binární řady denního výskytu/nevýskytu kategorií oblačnosti vytvořených z produktu CTY a ze staničních pozorování dle tabulky 2.3. Pro oba druhy měření máme tedy 3 binární řady denních výskytů pro každou oblast. Schopnost správně detekovat tyto kategorie byla testována pomocí kritérií představených v sekci 2.2.2.

Na základě výsledků produktu CTY můžeme odpovědět i na otázku, jak se mění vlastnosti produktu CFC v závislosti na vyskytující se typu oblačnosti, s nímž (díky námi vyděleným kategoriím) přímo souvisí i výška oblačnosti. Produkty CTY a CFC jsou, jak již bylo vysvětleno v teoretické části, úzce svázány, přičemž základní je produkt CFC a v podobě CTY se jedná o jeho jakousi podrobnější nadstavbu. Synoptická pozorování celkového pokrytí oblohy oblačností a výskytu jednotlivých typů spolu nejsou takto svázána. To nám ale nebrání v tom, abychom srovnali schopnost detekce produktem CTY s referenčním pozorováním výskytů, načež na základě tohoto srovnání můžeme usuzovat na vlastnosti produktu CFC. Může nám to napovědět, jak je jeho kvalita ovlivněna různou skladbou výskytu kategorií typu oblačnosti.

Zařazením tabulky 3.4 uvedené níže chceme dokumentovat obtíže vyvstávající při pokusu o validaci denního produktu CTY, oproti validaci produktu CFC. Jak již bylo napsáno, nejprve bylo nutné vytvořit vlastní speciální kategorie typů oblačnosti, čímž si vlastně z produktu CTY i z veličin  $C_1$ ,  $C_m$  a  $C_h$  vybíráme pouze omezenou informaci o výskytu jednotlivých pater oblačnosti. Zde pak vystupuje hlavní limitující faktor našeho studia, protože pozorovatel dokáže zaznamenat vyšší patra oblačnosti pouze, pokud nejsou zastíněna oblačností nižších pater. Družice sice dokáže nahlédnout do struktury vícevrstevnaté oblačnosti, avšak s množstvím překrývajících se oblaků samozřejmě klesá přesnost detekce.

Oblast	Počet dnů vstupujících do analýzy	Počet neanalyzovatelných dnů					
		Low		Middle		High	
		Stanice	CTY	Stanice	CTY	Stanice	CTY
Kocelovice	2283	14	93	296	93	920	93
Cheb	2283	15	78	376	78	1465	75
Kuchařovice	2283	56	75	784	75	1875	78
Liberec	2283	19	27	303	27	1239	23
Pardubice	2283	31	44	291	44	1260	40
Přerov	2283	23	43	372	43	1498	37
Sedlec	2283	34	113	308	113	1182	113
Tušimice	2283	26	73	659	73	1732	73
Mošnov	2283	10	41	283	41	1435	40
<b>Celkově:</b>	<b>20547</b>	<b>228</b>	<b>587</b>	<b>3672</b>	<b>587</b>	<b>12606</b>	<b>572</b>

*Tab. 3.4: Srovnání dostupnosti informací o jednotlivých kategoriích z družice a ze stanic.*

V tabulce 3.4 vidíme počty dnů, které musely být z analýzy vyjmuty buď kvůli neanalyzovatelným údajům ze stanice, či datům z družice MSG. Počet dnů vyřazených kvůli nekompletní informaci v produktu CTY je napříč kategoriemi téměř stejný, jen s lehkým poklesem u oblačnosti vyšších pater, kdy v pěti dnech byla družice schopna rozlišit pouze tento typ. Naproti tomu vidíme u staničních pozorování prudký vzestup neanalyzovatelných dnů s výškou oblačnosti. Rozdíly mezi jednotlivými oblastmi se nebudeme zabývat, protože mohou být důsledkem mnoha známých i neznámých příčin ovlivňujících pozorování na jednotlivých stanicích.

Ačkoliv k počtu neanalyzovatelných dnů kvůli produktu CTY musíme přičíst i 31 dnů, kdy jsou denní produkty CFC a CTY z MSG nedostupné (viz tabulka 1.3), ukazuje nám tab. 3.4 jednoznačný přínos produktu CTY v mnohem kompletnějších informacích o oblačnosti vyšších pater, než můžeme dostat ze staničních pozorování. Kvalitu těchto údajů pak můžeme za pomoci staničních pozorování jen těžko ohodnotit.

V tabulce 3.5 jsou uvedeny výsledky validačních kritérií souhrnně ze všech oblastí. Interpretace výsledků je následovná, jelikož je dnů s výskytem každé kategorie typu oblačnosti v její řadě většina, neliší se kritérium *TS* moc od *PC* a obě

dvě vyjadřují dobrou shodu řad. Ve všech třech kategoriích je jejich výskyt produktem CTY určen ve více než třech čtvrtinách případů stejně jako v referenční řadě. Dobrou schopnost produktu CTY rozpoznat a detekovat jednotlivé kategorie potvrzuje kritérium *B*. Kromě kategorie *high* produkt udává jen o trochu méně výskytů typů, než je tomu v referenční řadě.

*FAR*, pravděpodobnost špatné detekce výskytu, neboli podíl pozitivních detekcí MSG, které se ukázaly chybné, je velmi malý pro všechny kategorie, což opět naznačuje dobrou kvalitu produktu CTY. *H* ukazující v jaké části případů, kdy stanice detekovala typ oblačnosti, ho detekovala i MSG, dosahuje také vysokých hodnot ve všech kategoriích znamenající dobrý výsledek. Zatím tedy všechna kritéria naznačují dobrou shodu mezi CTY a referenčními řadami. Dle zatím diskutovaných kritérií jsou družicí o trochu hůře detekovány typy spadající do kategorie *middle*. Z dalších výsledků ale uvidíme, že je tomu přesně opačně.

*F* udává část dnů, v kterých se kategorie typu oblačnosti dle stanice nevyskytuje, ale je od MSG označena jako den s typem oblačnosti. Tento ukazatel je velký, což nepotvrzuje dobrou shodu mezi detekcemi. Zejména pro kategorie *low* a *high* jsme svědky překrytí 90 % dnů bez referenčního výskytu falešnou detekcí typů oblaků do nich spadajících od družice MSG. Heidkeho skóre potvrzuje právě objevený nesoulad mezi řadami naznačený kritériem *F*. Tedy že CTY se trefuje do výskytů oblačnosti dané kategorie pouze díky tomu, že ta se v drtivé většině případů vyskytuje, ale selhává při rozpoznávání situací, kdy se daná kategorie nevyskytuje. To dokládá i Pierceho skóre (neboli také Hanssen-Kuipers skóre), které hodnotí schopnost systému odlišit výskyt a nevýskyt. Lepší než náhodné určení výskytu/nevýskytu nacházíme pouze pro kategorii typu oblaků *middle*.

Pokusíme-li se najít rozdíly mezi kategoriemi, zjistíme, že schopnost detekce produktu CTY (bereme-li staniční pozorování jako popis reálného stavu atmosféry ve všech kategoriích) je sice stále špatná, ale nejlepší u kategorie *middle*. Bude to zřejmě tím, že detekce kategorie *low* je pro družici při zastínění vyššími patry velmi obtížná. Oproti tomu v kategorii *high* staniční pozorování braná jako referenční hodnoty ztlačují rozsah námi analyzovaného souboru a omezují ho jen na určité situace. To spolu s tím, že z geometrického hlediska (celkově to nemusí platit) je tato kategorie nejsnáze rozlišitelná družicí, vede k názoru, že informace o oblačnosti vyšších pater z obou zdrojů by se měly navzájem porovnávat jen ve speciálně ošetřených případech, jelikož je těžké říci, který je zatížen menší chybou.

<b>Kategorie</b>	<b>PC</b>	<b>TS</b>	<b>B</b>	<b>FAR</b>	<b>H</b>	<b>F</b>	<b>HSS</b>	<b>PSS</b>
<b>Low</b>	0,86	0,86	0,90	0,02	0,88	0,89	0,00	-0,01
<b>Middle</b>	0,77	0,76	0,85	0,06	0,80	0,67	0,06	0,13
<b>High</b>	0,88	0,87	0,99	0,06	0,93	0,92	0,01	0,01

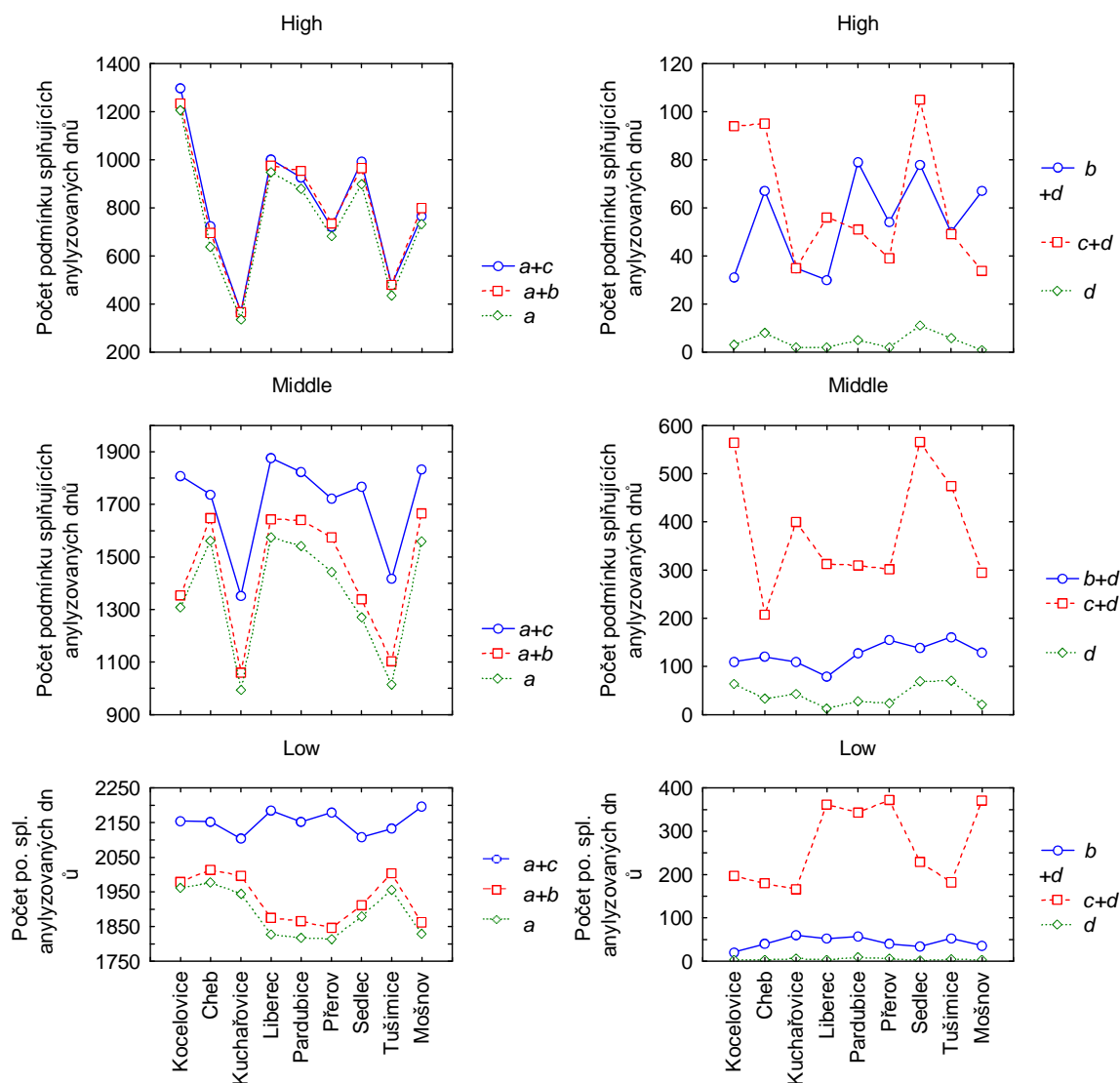
Tab. 3.5 : Celkové hodnoty validačních skóre.

V grafu 3.9 je demonstrována už poukazovaná slabina našeho validačního konceptu, kdy vlivem dostupných dat porovnáváme výskyty brané přes celý den, takže ve všech třech kategoriích je jen minimum případů, kdy se za celý den dle stanice neobjevil žádný typ oblaku spadající do těchto kategorií a CTY ho také nedetkovalo. Průběh počtu těchto případů v závislosti na oblasti můžeme sledovat v grafu 3.9 jako parametr  $d$ . Kategorie *middle* vykazuje nejvíce dnů bez výskytu typů oblaků do ní spadajících, ať již dle staničních pozorování, detekcí MSG nebo vzhledem k velikosti parametru  $d$  v této kategorii.

Dále si v levém sloupci grafu 3.9 můžeme povšimnout zajímavé geografické závislosti určení výskytu kategorie stanicí (modré body), CTY produktem (červené body) a oběma najednou (zelené body). V nepříbuzných (až na nadmořskou výšku) oblastech Kuchařovic i Tušimic detekují oba zdroje o mnoho méně dnů s výskytem oblaků spadajících do kategorií *middle* a *high*.

V oblastech Mošnova a Liberce, Pardubic a Přerova, které vycházely z validace produktu CFC nejlépe, vidíme v dolní části grafu 3.9 absolutně největší počet dnů napříč oblastmi, kdy stanice detekovala výskyt kategorie *low* a družice nikoliv. Jedná se tedy o počet dnů splňujících podmínku  $c$  z kontingenční tabulky, kterou můžeme v grafu vyčíst jako vzdálenost modrých a zelených bodů v levém sloupci a červených a zelených v pravém sloupci. V ostatních kategoriích přitom tyto stanice nejsou oproti ostatním takto výjimečné. Obě dvě uvedené geografické zvláštnosti ale nejsme schopni fyzikálně vysvětlit, nebo je dát do souvislosti v výsledky validace produktu CFC.



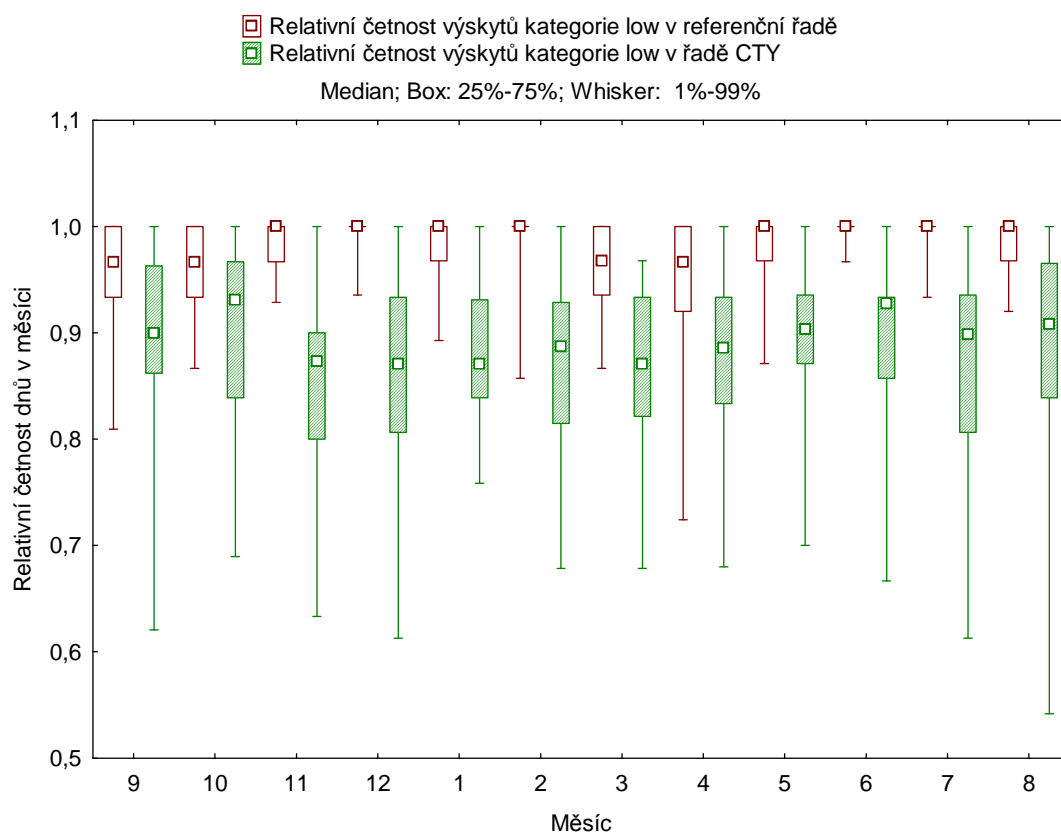


Graf 3.9: Počet členů kontingenční tabulky (viz obr. 2.3) v závislosti na kategorii *a* oblasti.

Nyní by nás zajímalo, jak se mění četnost dnů, v kterých se vyskytují jednotlivé kategorie oblačnosti, v závislosti na měsíci. Mohlo by nám to pomoci vysvětlit ne zcela dořešené problémy z podkapitoly 3.3, např. vývoj korelačního koeficientu  $r_p$ .

Vzhledem k tomu, že četnosti výskytů byly velmi ovlivněny počtem členů analyzované souhrnné řady ze všech oblastí v daném měsíci, uvádíme v níže uváděném grafu 3.10 relativní četnosti. Uvádíme vývoj relativních četností pouze pro kategorii *low*, jelikož v ostatních kategoriích je průběh velmi podobný. Roční chod je téměř nezatelný. Relativní četnost výskytu kategorie *low* v každém měsíci je v referenční řadě větší než devadesát procent pro výraznou většinu samostatných

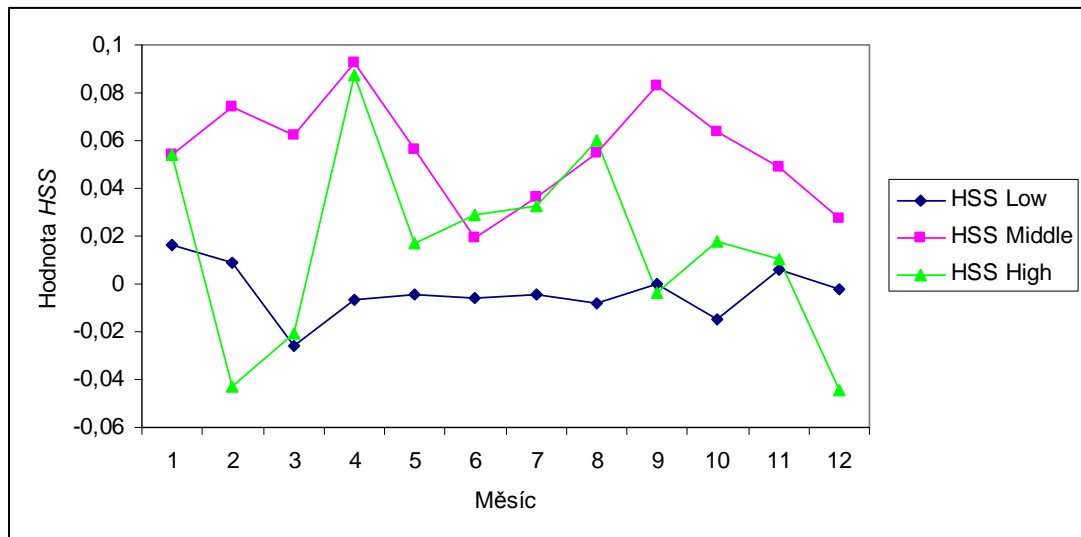
analýz měsíce, v řadě hodnot produktu CTY je relativní četnost výskytu *low* o něco menší.



Graf 3.10: Krabicový diagram relativní četnosti výskytu kategorie Low v závislosti na měsíci.

Roční chod kritéria *HSS*, jež porovnává schopnost produktu CTY udávat výskyt kategorií typů stejně jako denní souhrny staničních pozorování s náhodnou předpovědí výskytu/nevýskytu, je vykreslen v grafu 3.11. Nejedná se tentokrát o krabicový diagram, jelikož kvůli dostatečně velkému počtu členů kontingenční tabulky nemohlo být kritérium počítáno pro každý průběh měsíce. Byla tedy pro každý měsíc vytvořena jedna velká kontingenční tabulka a z ní následně počítáno Heidkeho skóre.

Ani tím jsme se však nevyhnuli problému našeho validačního konceptu, který spočívá v drtivé převaze detekovaných výskytů všech kategorií v obou řadách. To vede například k velice malým (i nulovým) hodnotám parametru *d*, což může rezultovat v záporné hodnoty *HSS*.



Graf 3.11: Roční chod Heidkeho kritéria.

Můžeme si v grafu 3.11 povšimnout rostoucí rozkolísanosti shodnosti detekce s výškou oblačnosti, což bude zapříčiněno právě proměnlivostí parametru  $d$ , který je nejmenší pro kategorii *low*. Těžko však z grafu 3.11 můžeme usuzovat na proměnlivou schopnost detekce výskytu/nevýskytu kategorií typu oblačnosti v závislosti na měsíci. Hodnota kritéria se totiž mění maximálně v řádu setin ideální dosažitelné hodnoty a je vždy blízká nule. Takže, i když např. červen vykazuje podobně jako v grafu 3.7 horší výsledky než červenec a srpen, vzhledem k malým rozdílům a diskutovaným limitům validace CTY nemůžeme jednoznačně tvrdit, že horší výsledky produktu CFC v tomto měsíci jsou zapříčiněny jinou skladbou typů oblaků a jejich odlišně přesným rozeznáváním.

Validace denního produktu CTY nám tedy žádné informace doplňující výzkum denního produktu CFC nepřinesla. Její výsledky navíc nejsou kvůli dostupným datům a použitému validačnímu konceptu dostatečně směrodatné, jelikož při porovnávání souhrnných denních výskytů dostaneme jen minimum dnů bez výskytu.

## Diskuse

Podstatná část této práce je věnována porovnání produktů CFC a CTY se synoptickými pozorováními příbuzných veličin. Použitý validační koncept spočíval v předpokladu, že přízemní pozorování jsou vyjádřením skutečného stavu atmosféry. Potom tedy veškeré charakteristiky shody řad jako systematická chyba, Pearsonův korelační koeficient a další, byly interpretovány jako schopnost (nebo odchylky od reality) družicových produktů udávat správný stav oblačnosti.

Výše popsaný postup nám dovolil snadnější popis a interpretaci výsledků, které ale většinou bylo obtížné fyzikálně zdůvodnit. Pokusme se nyní rozebrat situaci, kdy porovnáváme řady hodnot družicových produktů a staničních pozorování bez zjednodušujícího předpokladu staničních pozorování jakožto skutečného stavu atmosféry.

Okamžitá staniční pozorování jsou pak pouze odhadem skutečného stavu. Tento odhad je získán s nenulovou chybou a není ani v daném okamžiku naprosto reprezentativní pro celou oblast analýzy. I když totiž pozorovatel obsáhne svým pozorováním výseč oblohy o poloměru desítek kilometrů, tak to neznamená, že třeba o kilometr dál by pozorovaná hodnota nebyla jiná. Obsáhl by pak ve svém pozorování totiž i část oblohy dosud neviděnou. Naprosto reprezentativní odhad celkového pokrytí oblohy oblačností ( $N$ ) pro námi analyzovanou oblast (je určena polohou stanice a rozlišením družicového produktu) bychom dostali, jestliže bychom měli k dispozici pozorování veličiny  $N$  z každého bodu oblasti. V tomto odhadu veličiny  $N$  by ale byla zahrnuta i oblačnost sice z naší oblasti viděná, ale neležící v ní.

Družicový oblačný produkt CFC a CTY naopak vzniká na základě rozpoznání oblačnosti v jednotlivých pixelech snímku MSG a následným přeprojektováním na zemský povrch (spojeným s problémy s podpixelovou oblačností a správným určením výšky oblačnosti pro přesnou projekci na zakryté místo na zemském povrchu). Hodnotu CFC nebo CTY v naší oblasti pak získáme výpočtem z pixelů v ní ležících. Tímto způsobem dostaneme odhad pokrytí oblohy oblačností nebo zastoupení typů oblaků přímo nad studovanou oblastí. Tento odhad je samozřejmě také zatížený nenulovou chybou detekce a navíc ztrácíme kvůli omezenému rozlišení informaci z podpixelových měřítek.

Reuter a kol. (2009) porovnávali okamžitá synoptická pozorování s okamžitou informací MSG o oblačnosti v pixelu, kde stanice ležela, a poté také na pixelové úrovni denní průměry staničních pozorování s (pouze ze stejných časů jako pozorování vypočteným) denním CFC. Dospěli k mnohem lepší shodě řad než my v naší práci při hodnocení shody operativního CFC s denním průměrem veličiny N v oblastech velikosti gridboxu.

V rámci našeho validačního konceptu tento rozdíl přičítáme hlavně výběrové chybě denního CFC. Ta byla jednak způsobena rozdílností souboru použitého při tvorbě denního průměru CFC a toho, ze kterého se získává staniční průměr, ale také jsme do ní promítali problémy reprezentativnosti a prostorové kolokace.

Skutečně pravdivý odhad denního průměru pokrytí oblohy oblačností by měl být počítán jako integrál z kontinuálních celodenních měření všude na oblasti. Jak uvádí Reuter a kol. (2010), celková kvalita odhadu závisí na chybě každého okamžitého měření použitého pro průměrování, na korelaci mezi těmito jednotlivými měřeními, na přirozené prostoročasové variabilitě měřené veličiny a na její prostoročasové regularitě. Výběrová chyba jako část celkové chyby závisí hlavně na kombinaci posledních dvou jmenovaných. Záleží tedy zejména na fyzikální charakteristice měřené veličiny. Čím je přirozená variabilita veličiny větší, tím roste i počet potřebných okamžitých měření pro tvorbu důvěryhodného odhadu průměru.

Shrneme-li si dosavadní úvahy, tak při nezjednodušeném porovnání řad staničních průměrů a denního CFC bychom naráželi na problém ne zcela stejných veličin (pro CTY je to ještě markantnější), s tím spojený problém prostorové kolokace (také částečně přispívající do výběrové chyby) a problém různých celkových a v rámci nich i výběrových chyb členů obou druhů řad. Vzhledem k obtížně odhadnutelné variabilitě oblačnosti (advekce z okolí oblasti, vlastní časový vývoj) a kvůli tomu, že nebyl k dispozici kompletní soubor dat z MSG pro použití metody pro odhad výběrové chyby podle Reutera a kol. (2010), naráželi bychom v takovém nezjednodušeném porovnání na obou stranách (staniční i družicová data) na příliš mnoho neznámých. Zajímavé je, že samotná výběrová chyba oproti ideálnímu průměru je vzhledem k větší prostoročasové hustotě menší pro družicové oblačné produkty než pro staniční průměry.

Jak již bylo naznačeno, pro produkt CTY je úloha porovnání tohoto družicového produktu s pozemními měřeními (teoreticky i prakticky) obtížnější. Musíme si uvědomit, že klasické dělení oblačnosti na 10 základních druhů je jen

jedna z možných klasifikací oblaků. Je otázkou, nakolik je takováto klasifikace srovnatelná s klasifikací založenou na družicových datech. Je třeba si uvědomit, že zatímco jeden druh oblaku se může vyskytovat v různých výškách a je stále klasifikován stejně, jindy stačí malá změna výšky výskytu a podle klasifikace se pak například jedná buď o Altocumulus nebo už o Cirrocumulus. Z fyzikálního hlediska ale nemusí být v principu jejich vzniku podstatný rozdíl. Družicový pohled (a na něm založená klasifikace), který vychází z optické tloušťky oblaků, jejich teploty, mikrofyzikálního složení (kapičky, krystalky) a podobně, je z fyzikálního hlediska poněkud vhodnější než klasifikace pozorovatelská.

Námi provedená validace řad hodnot denních operativních produktů CFC a CTY vypovídá o jejich značné odlišnosti od referenčních řad. Tyto výsledky se ale nedají interpretovat tak, že analyzované produkty jsou málo kvalitní. Když jsme při popisu výsledků během práce používali slova jako „špatná“ a podobně, bylo to ve významu velmi odlišná. Z výsledků vyplývá, že pro účely klimatologie nemohou zatím tyto družicové produkty nahrazovat tradiční zdroje dat (je potřeba mnoha dalších srovnání), jejich přínos však spočívá v množství doplňujících informací o stavu atmosféry. Jedná se o informace často ze zemského povrchu principiálně nezískatelné (souvislé informace o vyšších patrech oblačnosti atd.), které navíc převyšují staniční pozorování svou prostorovou i časovou hustotou a homogenitou.

## Závěr

Nyní uvedeme krátký výtah z jednotlivých částí práce a shrneme výsledky našeho výzkumu. Přitom si budeme všímat, jak byly splněny cíle vytyčené v úvodu diplomové práce.

Byl podán stručný faktický přehled potřebný pro uvedení do zpracovávané problematiky. Byly představeny některé meteorologické družice a jejich hlavní senzory, organizace EUMETSAT, CM-SAF a jím poskytované družicové operativní oblačné produkty. Celou prací se pak prolínali citace současných výsledků výzkumu družicových oblačných produktů, z nichž byly přežaty i některé metody analýzy.

Dále byla popsána příprava vlastního výzkumu denních produktů CFC a CTY z MSG. Byly připraveny soubory dat a vypracován validační koncept pro jejich porovnání s korespondujícími denními průměry sestavenými ze staničních pozorování, která byla původně součástí části zprávy SYNOP. Sestavili jsme algoritmy potřebné pro přípravu dat v časové řady hodnot v analyzovaných oblastech, přičemž byly tyto řady očištěny od nehomogenit, aby byly vhodné k vzájemnému porovnávání. Byl zařazen důkladný popis dále v práci používaných statistických metod analýzy časových řad.

Přímá srovnání korespondujících si řad a jejich statistické charakteristiky ukázaly jejich velkou odlišnost a tedy i neurčitost produktu CFC větší než předpokládaných 10 %. Oproti studii Reutera a kol. (2009) jsme dospěli k mnohem větší chybě CFC a rozdělení diferencí jednotlivých členů od referenčních neodpovídá normálnímu rozdělení. Hlavním faktorem námi získaných větších chyb je pravděpodobně výběrová chyba, do které zahrnujeme odlišné průměrování od jejich práce i odlišnou prostorovou kolokaci družicového produktu a staničního průměru. Pro její odhad by však byla potřeba detailnějších vstupních dat, jak navrhuje Reuter a kol. (2010). Bylo zjištěno poměrně odlišné souhrnné rozdělení hodnot produktu CFC od referenčního rozdělení a předpokládaného gama rozdělení.

Zjistili jsme, že produkt CFC nedokáže plně vystihnout prostorovou variabilitu oblačnosti na území ČR. Bylo by však vhodné provést analýzu ještě vzhledem ke stanicím orograficky speciálnějším, např. horským. Nedokázali jsme potvrdit hypotézu o vlivu omezené frekvence pozorování na stanici na odlišnost řady CFC od referenční řady staničních průměrů (nicméně nejlepší výsledky vykazují řady CFC v lokalitách Pardubic a Přerova, tj. v místech s nejpravidelnějšími staničními

měřeními). K vylepšení shody řad nedošlo ani při analýze dnů s nejstacionárnější oblačností, kdy by měl být vliv výběrové chyby oproti ostatním dnům souboru minimální. Byl zjištěn nadhodnocený počet jasných dnů udávaných produktem CFC oproti referenčnímu výskytu (hlavně pro oblasti z Čech) a malá shoda řad v těchto dnech. Schopnost produktu pozorovat zamračené dny se ukazuje lepší.

Byl objeven roční chod systematické chyby, ovšem na rozdíl od práce Karlsson a Dybbroe (2010) je roční chod rozkolísaný s převažujícím záporným biasem v listopadu, červenci a srpnu místo největšího podceňování v zimních měsících. Dokonce se objevuje nadhodnocování množství oblačnosti produktem CFC v jarní části roku. Střední absolutní odchylka řad je napříč měsíci téměř konstantní. Chod Pearsonova koeficientu naznačuje, že shoda řady produktu CFC s referenční řadou je horší v zimních měsících a březnu, což může být způsobeno výskytem obtížněji pozorovatelných situací pro družici (např. tenké ciry nad promrzlým terénem). Záporné hodnoty  $r_p$  při některých průbězích měsíců naznačují klasicky opačně (menší/větší oproti průměru) udávané pokrytí oblohy oblačností od CFC a referenčních průměrů při některých oblačných situacích. Nejlepší je shoda v červenci, září a říjnu. Naopak podobnost celkového rozdělení hodnot se směrem od zimních měsíců k letním zhoršuje. Na shodu jednotlivých členů řad nemá překvapivě schopnost vystihnout správně celkové rozdělení hodnot vliv.

Závislost kvality produktu na výšce a druhu vyskytující se oblačnosti měla být zkoumána za pomoci validace produktu CTY. Ukázalo se, že v rámci našeho validačního konceptu, kdy porovnáváme řady denních výskytů s možnými hodnotami pouze ano/ne, nejsme schopni (vzhledem k drtivé většině dnů s výskytem kategorií typu oblaků) přenést získané informace na produkt CFC. Navíc schopnost produktu CTY správné detekce výskytu všech kategorií v daném dni je velmi slabá, nejméně špatná pro střední patro oblačnosti. Výsledky zkoumání prostorového a časového vývoje shodnosti CTY s referenčními výskyty nejsou dostatečně směrodatné, abychom z nich mohli učinit nějaké závěry. Hlavně validace pro kategorie *middle* a *high* je ovlivněna menším počtem členů analyzovaného souboru, jelikož jejich výskyt je ve velké části dnů ze stanice, kvůli zastínění nižšími patry oblačnosti, nepozorovatelný. Zde produkt CTY kvalitativně i kvantitativně převyšuje staniční pozorování. Vzhledem k inverzní pozorovací situaci té právě diskutované je pro produkt CTY nejtěžší detekovat kategorii *low*. Celkově je validace produktu



CTY obtížnější, jelikož už samotná detekce typu oblačnosti (ať už pozorovatelem nebo družicí) je mnohem těžší a je zatížena velkou neurčitostí.

## Seznam použitých zkratek

Ac	Alto cumulus
AMSU	Advanced Microwave Sounding Unit
As	Altostratus
ATOVS	Advanced TIROS Operational Vertical Sounder
ATS	Applications Technology Satellite
AVHRR	Advanced Very High Resolution Radiometer
A-train	Afternoon-train
<i>B</i>	Vychýlení (bias)
CALIOP	Cloud-Aerosol Lidar with Orthogonal Polarization
Cb	Cumulonimbus
Cc	Cirrocumulus
CDOP	Continuous Development and Operations Phase
CDR	Climate Data Records
CEOS	Committee on Earth Observation Satellites
CERES	The Clouds and the Earth's Radiant Energy System
CFC	Cloud Fractional Cover
C <sub>h</sub>	Oblaky druhů: Ci, Cc, Cs
Ci	Cirrus
C <sub>l</sub>	Oblaky druhů: St, Sc, Cu, Cb
C <sub>m</sub>	Oblaky druhů: Ac, As, Ns
CMa	Cloud Mask
CM-SAF	SAF on Climate Monitoring
COT	Cloud Optical Thickness
CPH	Cloud Phase
CPR	Cloud Profiling Radar
Cs	Cirrostratus
CTH	Cloud Top Height
CTP	Cloud Top Pressure
CTT	Cloud Top Temperature
CTY	Cloud Type
Cu	Cumulus
CWP	Cloud Water Path

ČHMÚ	Český hydrometeorologický ústav
<i>D</i>	Detekováno (družice)
DMSF	Defense Meteorological Satellites Program
DWD	Deutscher Wetterdienst
ESA	European Space Agency
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
<i>F</i>	False alarm rate
<i>FAR</i>	False alarm ratio
FCDR	Fundamental Climate Data Records
GCOS	Global Climate Observing System
GEO	Group on Earth Observations
GERB	Geostationary Earth Radiation Budget
GPS	Global Positioning System
GRAS	Global Navigation Satellite System Receiver for Atmospheric Sounding
GRAS-SAF	SAF on GRAS Meteorology
<i>H</i>	Hit rate
HIRS	High resolution Infrared Radiation Sounder
HLW	Layered precipitable water
HRV	High resolution visible
HSH	Temperature and specific humidity at 6 pressure levels
<i>HSS</i>	Heidke Skill Score
HTW	Total precipitable water
H-SAF	SAF on Support to Operational Hydrology and Water Management
IAA	Inner Arctic Circle
IASI	Infrared Atmospheric Sounding Interferometer
IBA	Initial Baseline Area
ICSU	International Council for Science
<i>ID</i>	Mezikvartilová vzdálenost
IOC	Intergovernmental Oceanographic Commission
IOP	Initial Operations Phase
IR	Infrared

LSA-SAF	SAF on Land Surface Analysis
<i>MAE</i>	Střední absolutní odchylka
<i>ME</i>	Systematická chyba
<i>MED</i>	Podmíněný medián
MHS	Microwave Humidity Sounder
<i>MO</i>	Průměrná hodnota
MODIS	Moderate Resolution Imaging Spectroradiometr
<i>MSE</i>	Střední kvadratická chyba
MSG	Meteosat druhé generace
<i>N</i>	Počet prvků
N	Celkové pokrytí oblohy oblačností
NASA	The National Aeronautics and Space Administration
NIR	Near infrared
NOAA	The National Oceanic and Atmospheric Administration
Ns	Nimbostratus
NWC-SAF	SAF on Support to Nowcasting and Very Short Range Forecasting
NWP-SAF	SAF on Numerical Weather Prediction
OSI-SAF	SAF on Ocean and Sea Ice
OSN	Organizace spojených národů
O3M-SAF	SAF on Ozone & Atmospheric Chemistry Monitoring
<i>P</i>	Pozorováno (stanice)
<i>PC</i>	Proportion correct
POES	Polar Operational Environmental Satellite
<i>PSS</i>	Pierceho skóre
<i>RMSE</i>	Odmocnina ze střední kvadratické chyby
<i>r<sub>p</sub></i>	Pearsonův korelační koeficient
SAF	Satellite Application Facility
SAL	Surface albedo
Sc	Stratocumulus
SDL	Downward long-wave radiation
SEVIRI	Spinning Enhanced Visible and Infrared Imager
SIS	Incoming short-wave radiation
SNL	Net long-wave radiation

SNS	Net short-wave radiation
SOL	Outgoing long-wave radiation
SRB	Surface radiation budget
SSD	Směrodatná odchylka
SSM/I	Special Sensor Microwave/ Imager
St	Stratus
SYNOP	Surface synoptic observations
SVAR	Rozptyl
TCDR	Thematic Climate Data Records
TET	Emitted thermal radiative flux
TIROS	Television Infrared Observation Satellite
TIS	Incoming solar radiative flux
TOA	The top of the atmosphere
TRS	Reflected solar radiative flux
TS	Threat score
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNEP	The United Nations Environment Programme
UNFCC	United Nations Framework Convention on Climate Change
USA	The United States of America
UV	Ultraviolet
VIS	Visible
WMO	The World Meteorological Organization
X	Obecná hodnota

## Literatura a použité internetové odkazy

Capderou, Michel, 2005: Satellites. orbits and missions. *Springer-Verlag France*, Paris.564 p., il.

Derrien, M., and H. LeGleau, 2005: MSG/SEVIRI cloud mask and type from SAFNWC. *Int. J. Rem. Sens.*, **26**, 4707-4732.

Dürr, Bruno , Zelenka, Antoine , Müller, Richard and Philipona, Rolf, 2010: Verification of CM-SAF and MeteoSwiss satellite based retrievals of surface shortwave irradiance over the Alpine region, *International Journal of Remote Sensing*, **31**, 4179-4198.

Dybbroe, A., A. Thoss and K.-G. Karlsson, 2005a: SAFNWC AVHRR cloud detection and analysis using dynamic thresholds and radiative transfer modelling – Part II: Validation. *J. Appl. Meteor.* **44**, 55-71.

Dybbroe, A., A. Thoss and K.-G... Karlsson, 2005b: SAFNWC AVHRR cloud detection and analysis using dynamic thresholds and radiative transfer modelling – Part I: Algorithm description. *J. Appl. Meteor*, **44**, 39-54.

Journée, M., C. Bertrand, 2010: Improving the spatio-temporal distribution of surface solar radiation data by merging ground and satellite measurements. *Remote Sens. Environ.*, **114**, pp 2692-2704.

Karlsson, K.-G., U. Willén, C. Jones, and K. Wyser, 2008: Evaluation of regional cloud climate simulations over Scandinavia using a 10-year NOAA Advanced Very High Resolution Radiometer cloud climatology, *J. Geophys. Res.*, **113**, D01203, doi:10.1029/2007JD008658.

R. R. Kelkar, 2007: Satellite Meteorology, *BS Publications* 251 s, ISBN 8178001373.

Macke, A.; Kalisch, J.; Hollmann, R., 2010: Validation of downward surface radiation derived from MSG data by in-situ observations over the Atlantic ocean, *Meteorologische Zeitschrift Vol. 19*, No. 2 (2010), p. 155 – 167, DOI: 10.1127/0941-2948/2010/0433

Meteorologický slovník (1993): kolektiv autorů: Meteorologický slovník výkladový a terminologický. MŽP, Praha, 594 s.

M. Reuter, W. Thomas, P. Albert, M. Lockhoff, R. Weber, K. G. Karlsson and J. Fischer, 2009: The CM-SAF and FUB cloud detection schemes for SEVIRI: Validation with synoptic data and initial comparison with MODIS and CALIPSO, *J. Appl. Meteorol. Clim.*, **48**, no. 2, pp. 301–316.

Reuter, M., Thomas, W., Mieruch, S., and Hollmann, R., 2010: A Method for Estimating the Sampling Error Applied to CMSAF Monthly Mean Cloud Fractional Cover Data Retrieved From MSG SEVIRI, *IEEE T. Geosci. Remote*, **48**, 2469–2481, doi:10.1109/TGRS.2010.2041240.

Roebeling, R. A., A. J. Feijt, and P. Stammes, 2006: Cloud property retrievals for climate monitoring: Implications of differences between Spinning Enhanced Visible and Infrared Imager (SEVIRI) on METEOSAT-8 and Advanced Very High Resolution Radiometer (AVHRR) on NOAA-17, *J. Geophys. Res.*, **111**, D20210, doi:10.1029/2005JD006990.

Schulz, J.; Albert, P.; Behr, H.-D.; Caprion, D.; Deneke, H.; Dewitte, S.; Duerr, B.; Fuchs, P.; Gratzki, A.; Hechler, P.; Hollmann, R.; Johnston, S.; Karlsson, K.-G.; Manninen, T.; Mueller, R.; Reuter, M.; Riihelae, A.; Roebeling, R.; Selbach, N.; Tetzlaff, A.; Thomas, A.; Werscheck, M.; Wolters, E.; Zelenka, A., 2009: Operational climate monitoring from space: the EUMETSAT Satellite Application Facility on Climate Monitoring (CM-SAF), *Atmos. Chem. Phys.*, **9**, Pages: 1687-1709, doi: 10.5194/acp-9-1687-2009

Radim Tolasz et al., 2007: Atlas podnebí Česka [kartografický dokument]. 1. vyd. Praha : Český hydrometeorologický ústav ; Olomouc : Univerzita Palackého v

Olomouci, 255 s, il., tab., mapy. ISBN 9788086690261 (ČCHMU). ISBN 9788024416267 (UP)

Träger-Chatterjee, Christine; Müller, Richard W.; Trentmann, Jörg; Bendix, Jörg; 2010: Evaluation of ERA-40 and ERA-interim re-analysis incoming surface shortwave radiation datasets with mesoscale remote sensing data, *Meteorologische Zeitschrift*, **19**, issue 6, pp. 631-640.

Willén, U., 2008: Preliminary use of CM SAF cloud and radiation products for evaluation of regional climate simulations, *SMHI Reports Meteorology and Climatology*, 131, 48 pp.

World Meteorological Organization, 1956: International Cloud Atlas: Abridged atlas. *World Meteorological Organization*, 62 pp. and 72 plates.

CM-SAF, 2011: Annual Product Quality Assessment Report 2010, Reference number SAF/CM/DWD/VAL/OR6, Version 1.1, 105 pp. (Available online at [http://www.cmsaf.eu/bvbw/generator/CMSAF/Content/Publication/pdf/SAF\\_\\_CM\\_\\_DWD\\_\\_VAL\\_\\_OR6\\_\\_1\\_\\_1,templateId=raw,property=publicationFile.pdf/SAF\\_CM\\_DWD\\_VAL\\_OR6\\_1\\_1.pdf](http://www.cmsaf.eu/bvbw/generator/CMSAF/Content/Publication/pdf/SAF__CM__DWD__VAL__OR6__1__1,templateId=raw,property=publicationFile.pdf/SAF_CM_DWD_VAL_OR6_1_1.pdf))

Fuchs P. et al., 2009: CDOP Service Specifications, Reference number SAF/CM/DWD/SeSp/1, Version 1.5, 35 pp. (Available online at [http://www.cmsaf.eu/bvbw/generator/CMSAF/Content/Publication/cdop\\_\\_service\\_\\_specifications\\_\\_1.5,templateId=raw,property=publicationFile.pdf/cdop\\_service\\_specifications\\_1.pdf](http://www.cmsaf.eu/bvbw/generator/CMSAF/Content/Publication/cdop__service__specifications__1.5,templateId=raw,property=publicationFile.pdf/cdop_service_specifications_1.pdf))

Karlsson K. G., Meirink J. F., Stengel M., 2011: Product User Manual - CLOUDS, Reference number SAF/CM/ DWD/PUM/CLOUDS, Version 1.5, Ed. Hollmann R., CM-SAF, 95 pp. (Available online at [http://www.cmsaf.eu/bvbw/generator/CMSAF/Content/Publication/SAF\\_\\_CM\\_\\_DWD\\_\\_PUM\\_\\_CLOUDS\\_\\_1.5,templateId=raw,property=publicationFile.pdf/SAF\\_CM\\_DWD\\_PUM\\_CLOUDS\\_1.pdf](http://www.cmsaf.eu/bvbw/generator/CMSAF/Content/Publication/SAF__CM__DWD__PUM__CLOUDS__1.5,templateId=raw,property=publicationFile.pdf/SAF_CM_DWD_PUM_CLOUDS_1.pdf))



WMO TD No. 1485, 2009: Recommendations for the Verification and Intercomparison of QPFs and PQPFs from Operational NWP Models – Revision 2 - October 2008, World Meteorological Organization, 43 pp. (Available online at [http://www.wmo.int/pages/prog/arep/wwrp/new/documents/WWRP2009-1\\_web\\_CD.pdf](http://www.wmo.int/pages/prog/arep/wwrp/new/documents/WWRP2009-1_web_CD.pdf))