

VYSOKO DOSTUPNÁ DATABÁZA DIAĽKOVÉHO PRIESKUMU ZEME

Ján KIANIČKA

GeoModel Solar s.r.o., Pionierska 15, 831 02, Bratislava, Slovenská Republika
jan.kianicka@geomodel.eu

Abstrakt

Práca s digitálnymi geografickými dátami väčšinou predstavuje manipuláciu veľkých objemov dát v rôznych formátoch. Je potrebné riešiť ich bezpečné ukladanie, jednoduchý a rýchly prístup k nim v správnych časových a priestorových súradniciach. Dáta diaľkového prieskumu zeme (DPZ) v digitálnej forme predstavujú dvoj-dimenzionálne rastre v určitej mierke a rozlíšení. Plus, pre ďalšie spracovanie sú k dispozícii navigačné a kalibračné údaje. Pre klimatologické štúdie je potrebné mať k dispozícii dlhé časové série týchto dát. Keď berieme do úvahy plné priestorové a časové rozlíšenie geostacionárnych meteorologických satelitov, pracujeme s objemami dát rádovo v 10-tkach TB, dátové manipulácie sú naozaj obtiažne.

Tento článok pojednáva o tom ako sme vybudovali robustnú vysoko dostupnú databázu dát diaľkového prieskumu zeme. Spracovali sme dáta z rôznych misií a senzorov: Meteosat prvej aj druhej generácie, americký GOES a japonský MTSAT. Tým pádom je pokrytý takmer celý povrch Zemský. Naše požiadavky na databázu boli hlavne jednoduchý a jednotný prístup k dátam, ukladanie dát vo vhodnej kompresii, transformovanie z špecifických dátových formátov do jednotného formátu, pokrytie služieb poskytujúcich dáta, vyriešenie rôznej projekcie a priestorového rozlíšenia, ako aj vyriešenie časovo-priestorových prekryvov. Článok obsahuje prehľad senzorov a ich špecifik, popis originálnych formátov, nasleduje prezentácia samotného dátového spracovania spolu s podporným softvérovým vybavením. Riešenie je založené prevažne na open-source platforme pokryté hlavne operačným systémom Linux, C/C++ nástrojmi na čítanie binárnych dát, plus integračná platforma je python. Na ukladanie rastrových dát bol použitý formát netCDF4 ktorý je založený na HDF5 formáte (Hierarchical Data Format). Pomocné navigačné, kalibračné a riadiace dáta sú ukladané v špeciálne dizajnovanej relačnej databáze. Pôvodné surové dáta z geostacionárnych senzorov zaberajú viac ako 100 TB diskového priestoru. Uložené vo vhodne nadizajnovaných netCDF4 súboroch zaberajú len niekoľko desiatok TB. Obalené špecializovaným softvérom a podpornou relačnou databázou je možné rýchle načítanie časového radu pre jeden bod ako aj čítanie časovo-priestorových výrezov. Pomocou kalibrácie je možné ľahko vypočítať odvodené fyzikálne veličiny ako napr. albedo alebo teplotu pri infračervenom spektre. V našich aplikáciách sa surové satelitné dáta využívajú v kombinácii s inými meteorologickými produktmi na odvodenie slnečnej radiácie dopadajúcej na zemský povrch. Technologické riešenie je ale možné použiť aj iných podobných aplikáciách DPZ. Infraštruktúra slúži ako solídny backend webového portálu SolarGIS.

Abstract

Work with geographical information in digital form relies on ability to process huge amount of data from various sources. Key points are safe storage, simple and fast access using geographical position and time coordinates. Remote sensing data presents multi-dimensional digital grids in a certain spatial resolution and spatial extent. In case of further processing, calibration and navigation information is available. For provision of climatological studies, long term time series of remote sensing data are needed. Considering full resolution, all required spectral channels and several years of geostationary meteorological satellite data, their manipulation constitutes a challenging problem.

This article presents processing of raw satellite data into a robust, versatile remote sensing database. We processed data from various platforms, Meteosat First Generation, Meteosat Second Generation, GOES and

MTSAT missions covering whole Earth's globe to build up a harmonized database for further usage in meteorological/solar radiation products development. The key requirements are: unified access to data, fast access in whatever dimension, storage in reasonable compressed format, transformed from different original mostly binary formats, in variable projections, in different time-spatial and spectral resolution, containing overlaps in regional and time domain. To fulfill these needs a robust remote sensing database was designed and implemented. At the beginning description of original formats with their specialities is briefly presented, main steps of the processing and programming supporting tools are presented as well. It is based on open source free accessible platforms covered by Linux operating system, various readers in C/C++, and python environment. For storing remote sensing data netCDF4 format was used. This is based on HDF5 format (Hierarchical Data Format) which is designed for multidimensional data handling. Auxiliary data and supporting metadata were stored in a relational database with high-performance tables design. All above-mentioned satellite data in their original binary formats might require more than 100 TB disk space. Stored in well designed compressed netCDF files it is a few TB. Wrapped in carefully tailored software tools and supported by metadata, it allows fast access to time series of data for one point, retrieving space-time slices, calibrate raw digital counts to physical units as true radiance, albedo or temperature for infrared spectral bands, apply additional geometrical corrections. In further processing it is used in combination with other meteorological data to derive various solar irradiation products. The database provides a firm raw data background for SolarGIS portal. However, this kind of approach may be used for other remote sensing applications where high data volumes are being processed, even in higher scales and smaller spatial-time extent.

Kľúčové slová: geografické dáta, diaľkový prieskum zeme, návrh databázy

Keywords: geographical data, remote sensing, database design

1 ÚVOD

Satelitné dáta v dnešnej dobe predstavujú cenný zdroj informácií o zemskom povrchu, ktoré vhodne dopĺňajú pozemné merania. Operatívne využívanie obrazových dát z geostacionárnych meteorologických satelitov zabezpečuje predpoveď počasia pre celý svet. Meteorologické satelity na sledovanie atmosféry majú dnes už viac ako 30 ročnú históriu a najnovšie senzory majú vysoké geometrické, spektrálne aj časové rozlíšenie. Tiež je možné v celku jednoducho tieto dáta získavať v reálnom čase, či už prostredníctvom priameho príjmu cez geostacionárne telekomunikačné satelity, alebo internetovým prenosom z dátových archívov. Prevádzkové centrá našťastie poznajú hodnotu aj starých archívnych dát a poskytujú aj dlhoročný dátový archív. Európsky Eumetsat a americká NOAA poskytujú online službu na sťahovanie archívnych dát v rôznych formátoch [1,2]. Dáta meteorologických satelitov majú relatívne nízke priestorové rozlíšenie s veľkosťou pixela 1-5 km v nadire. Ich pridaná hodnota je v globálnom pokrytí, častej frekvencii snímania a stabilnej rádiometrii senzora. Pre dlhodobé klimatické štúdie sú to prakticky jediné použiteľné obrazové dáta diaľkového prieskumu zeme.

2 PREHĽAD VYUŽÍVANÝCH METEOROLOGICKÝCH SATELITOV

Meteorologické geostacionárne satelity sa začali vypúšťať v 70-tych rokoch minulého storočia. V západnej Európe to boli Meteosaty dizajnované v ESA (European Space Agency). Meteosaty prvej generácie 1 až 7 pokrývajú kontinuálne celé obdobie od roku 1977 a majú len veľmi malé dizajnové rozdiely. Sú tzv. „*spin stabilized*“, teda stabilizované rýchlou rotáciou celého tela satelitu, u MFG až 100 otáčok za minútu. MFG 6 a 7 sú ešte stále funkčné na orbite a zabezpečujú *Indian Ocean Data Coverage* (IODC). Otáčaním satelitu a bodovým sensorom so spektrálnymi zrkadlami je zosnímaný jeden riadok obrazu. Vertikálnym pohyblivým zrkadlom je zabezpečený riadkový posun a takto je vytvorený celý disk. Na nultom poludníku sú už od roku 2004 Meteosaty druhej generácie [3,4]. Majú podobné riešenie senzor aj stabilizáciu na orbite ako MFG, ale

poskytujú až 12 spektrálnych kanálov, plus doplnkový viditeľný spektrálny kanál s vysokým geometrickým rozlíšením, až 1km x 1km. Geometrická a rádiometrická stabilita je vyššia ako u starších satelitov.

Druhým svetovým centrom vývoja takýchto satelitov je americká NASA. Tu tiež v 70-tych rokoch začal vývoj geostacionárnych satelitov radu GOES (Geostationary Operational Environmental Satellite). Od roku 1975 boli postupne vypúšťané satelity GOES 1 až 15, ktoré mali rôznu konštrukciu a tiež dosť rozdielne senzory a obrazové dáta [5]. GOES 1 až 7 boli podobné ako Meteosaty „*spin-stabilized*“, čomu zodpovedala aj konštrukcia senzora. Od GOES 8 už boli satelity konštruované ako tzv. „*three axes stabilized*“ a boli vybavené dvoma rôznymi zobrazovacími senzormi: Imager (Visible and Infrared Spin Scan Radiometer) a Sounder (VISSR Atmospheric Sounder) [6]. Ide o senzory s trochu vyšším geometrickým aj rádiometrickým rozlíšením, pixel má 0,5 – 1km vo viditeľnom spektre, a hlavne konštrukcia satelitu aj senzorov umožňuje snímanie akéhokoľvek výrezu v rôznom časovom intervale, čo umožňuje napr. sledovanie vývoja hurikánu v 5 minútových intervaloch. Pre budovanie globálnej klimateckej databázy sú takéto dáta problematickejšie ako tie zo „*spin-stabilized*“ satelitov.

Tretím centrom vývoja meteorologických satelitov bolo Japonsko (JMS – Japan Space Agency) v spolupráci s NASA tiež od 70tych rokov konštruuje a vypúšťa satelity radu GMS (Geostationary Meteorological Satellite). Od roku 1977 boli postupne vypustené GMS 1 až GMS 6 s veľmi podobným dizajnom ako prvé GOES satelity - „*spin-stabilized*“ [7]. GMS satelity pokrývali západný pacifik. Tie boli následne nahradené novšími „*three axes stabilized*“ MTSAT satelitmi (Multifunctional Transport Satellites) konštrukčne podobné novým GOES-om.

Všeobecne platí, čím starší je senzor, tým obtiažnejšie je dáta získať a využiť ich. Pre naše účely postačuje 10-20 rokov dát. Toto určuje hranicu do akej minulosti potrebujeme v dátovom spracovaní ísť. Samozrejme, čím dlhšie je obdobie, tým reprezentatívnejšie sú výsledky. V Tab.1 je prehľad ktoré dáta boli spracované.

Tab 1. Pre 0-tý poludník bolo spracovaných až 18 rokov dát. Pre Americký kontinent to je zatiaľ iba 12 rokov z GOES EAST pozície. Pacifická pozícia je zatiaľ iba v prípravnej fáze. Tieto štyri hlavné misie pokrývajú celý zemský povrch okrem polárnych oblastí.

Rok	0-tý poludník	IODC (57,5° E)	GOES EAST (75° W)	Pacific (145° E)
2011	MSG 1,2	MFG 7	GOES 8, 12, 13	MTSAT 1,2
2010				
2009				
2008				
2007				
2006				
2005	MFG 3-7	MFG 5		GOES 9
2004				
2003				
2002				
2001				
2000				
1999				
1998				
1997				
1996				
1995				
1994				

MFG – 3 spektrálne kanály, geometrické rozlíšenie 2,5 a 5 km

MSG – 12 spektrálnych kanálov, geometrické rozlíšenie 3km

GOES – 5 spektrálnych kanálov, geometrické rozlíšenie 1 a 4 km

MTSAT – 5 spektrálnych kanálov, geometrické rozlíšenie 1 a 4 km

GMS – 3 spektrálne kanály, geometrické rozlíšenie 1,2 a 5 km

3 DÁTOVÉ FORMÁTY

Satelitné obrazové dáta je možné získať v rôznych formátoch, od binárnych interných formátov až po jednoduché rastrové obrázky. Väčšinou platí, že odvodené dátové formáty majú jednoduchšie čítanie ale často nepreberajú z pôvodných formátov podporné hlavičkové informácie o stave senzora, geometrii a kalibrácii. Z tohto dôvodu používame binárne zdrojové dáta v špecifických formátoch, ktoré čítame buď pomocou poskytnutých softvérových modulov alebo implementujeme čítanie podľa dostupnej dokumentácie. Okrem samotných obrazových dát z hlavičky preberáme kalibračné údaje a pozíciu satelitu (tvz. Sub-satellite point).

Originálne formáty satelitných dát:

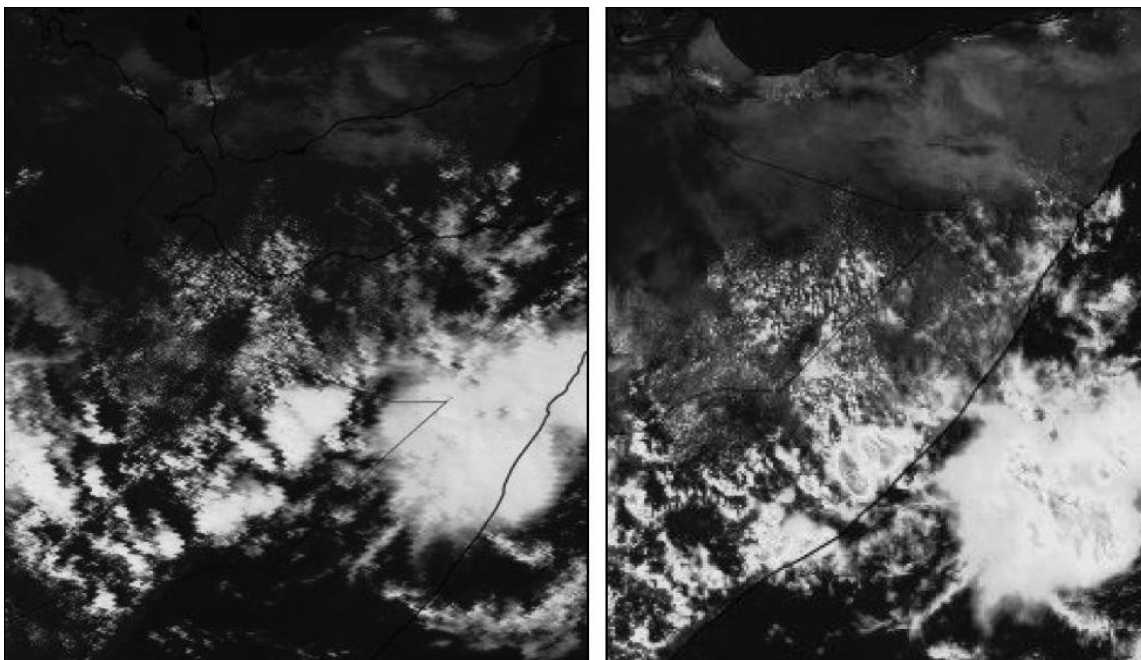
MSG dáta: HRIT formát (High Rate Information Transmission) [8] Ide o univerzálny formát na prenos obrazových dát diaľkového prieskumu zeme medzi systémami. Je rozdelený na bloky s vyčlenenými metadátami. Umožňuje aj kompresiu dát.

MFG dáta: OpenMTP (Open Meteosat Transition Programme) [9] Tiež ide o binárny obrazový formát skladajúci sa z hlavičky a obrazovej číselnej matice. Pre každý spektrálny kanál existuje jeden súbor a spojené sú pomocou jednoduchej „zip“ kompresie.

GOES dáta: GVAR (Goes Variable Format) [6] Ide tiež o špeciálny binárny formát štruktúrou zodpovedajúci prenosu dát zo satelitu 10 bitovo kódovaný. Všetky spektrálne kanály sú spojené a každý riadok je vybavený metadátovou hlavičkou. Výhoda je, že napriek vysokému rozlíšeniu viditeľného spektra je dátový objem v celku malý.

MTSAT dáta: Rovnako ako MSG ide o HRIT.

GOES 9 a GMS dáta: Tieto dáta sme obdržali v už odvodenom formáte, McIDAS Area (Man Computer Interactive Data Access systém Area) [10]. Ide tiež o binárny špecializovaný formát vyvinutý univerzitou vo Wisconsin. Originálne dáta z GMS satelitov už v dnešnej dobe nie sú prakticky čitateľné a nikto neposkytuje verejne dostupný softvér na ich spracovanie.



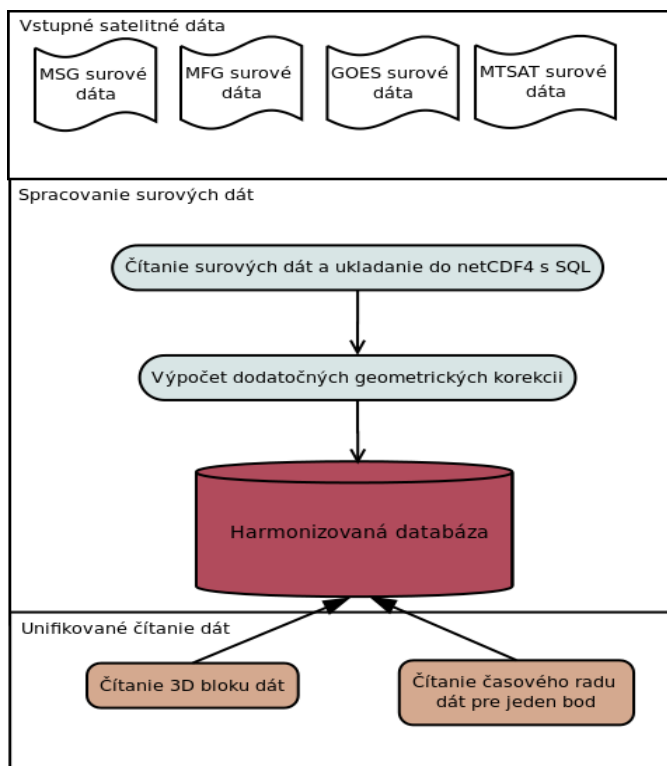
Obr.1. V satelitných centrách prebieha predspracovanie obnášajúce geometrické korigovanie, normalizáciu, „destripping“. Používame teda už korigované dáta. Na príklade je MFG snímka Somálskeho polostrova, pred „predspracovaním“ v ľavo, po „predspracovaní“ v pravo preložená hranicami kontinentu a štátov. Vidíme výrazný posun v polohe a „stripping“ na oblakoch.

Otázka znie, aké sú objemy týchto dát a ako postupovať pri ich spracovaní. Pre sedem rokov MSG dát, kde jedna snímka celého disku sa zaznamená každých 15 minút, a iba pre vybraných päť spektrálnych kanálov, to predstavuje približne 5.9 TB v komprimovanom pôvodnom HRIT formáte. MFG surové dáta pre 0-tý poludník plus IODC pozíciu to predstavuje 4.5 TB dát v komprimovanom tvare, kde každú pol hodinu sú zosnímané tri spektrálne kanály. GOES satelity snímajú dáta v rôznych priestorových výrezoch a v rôznych časových krokoch, ale každé tri hodiny snímajú celý disk. Dvanásť rokov trojhodinových dát v GVAR formáte zaberá 14 TB. Dáta pokrývajúce Pacifik, MTSAT, GOES9 a GMS pre 12 ročné obdobie v surovom komprimovanom formáte zaberajú 26 TB. Keď tieto čiastkové objemy sčítame, máme 50 TB surových dát v rôznych binárnych formátoch, uložených na veľkokapacitných páskach a na externých diskoch. Väčšinou sú komprimované a spektrálne kanály navzájom spojené v archívnych súboroch. S takto uloženými dátami nie je možné rýchle načítanie časového radu meraní pre lokalitu alebo vybraný región. Dáta musia byť rýchlo dostupné, geometricky korektné, kalibrované.

4 VYTVORENIE HARMONIZOVANEJ DATABÁZY

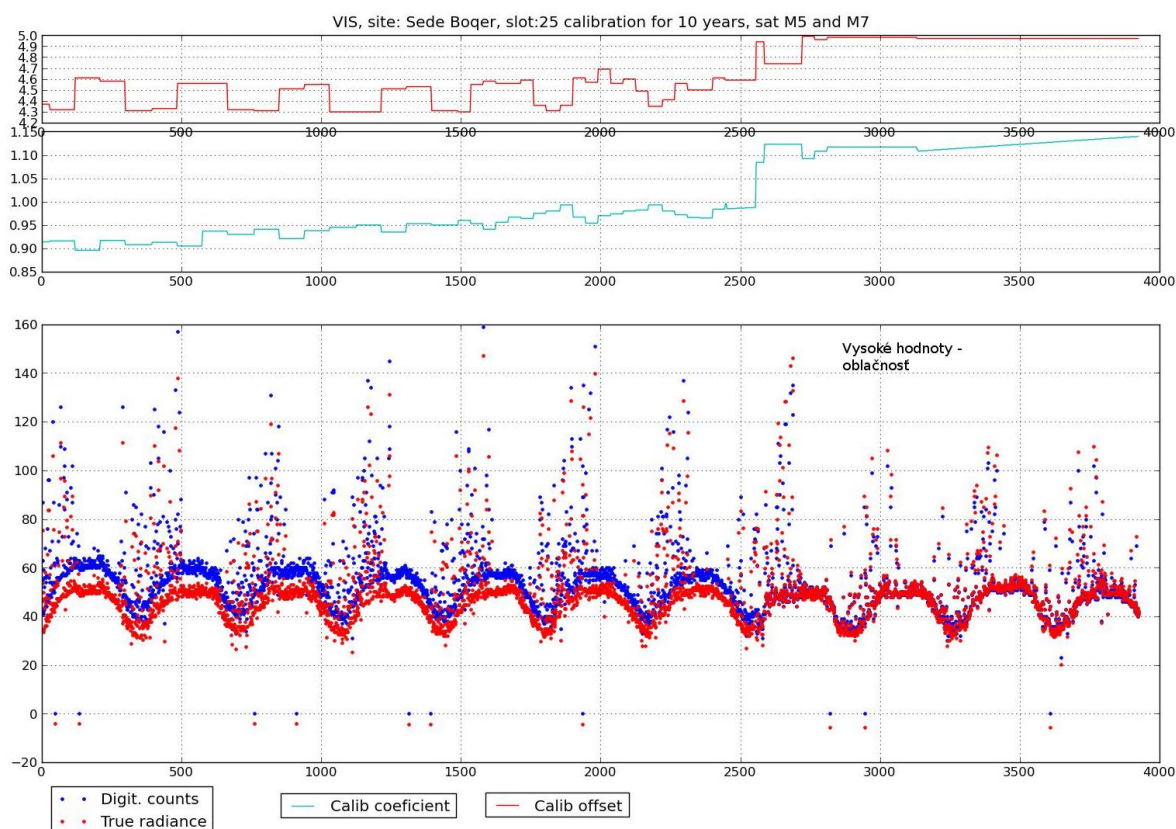
Požiadavky sú nasledovné, rýchly a unifikovaný prístup pre dáta z ktoréhokoľvek senzora pre ľubovoľné časové obdobie. Musí byť možnosť previesť originálne strojové jednotky (digital counts) do fyzikálnych veličín akými sú napríklad radiácia pre viditeľné spektrum, odvodená teplota pre infračervené spektrum. Dáta musia byť navzájom interkalibrované a musia zohľadniť aj stratu citlivosti senzora vo voditeľnom spektre v čase. Všetky snímky musia byť čo najpresnejšie navzájom geometricky zosúladené aj v prípade, že došlo ku

pohybu satelitu alebo snímaniu z inej pozície, prípadne samotný obraz mal v sebe distorziu. Čo je najdôležitejšie, musí byť možné databázu jednoducho aktualizovať o nové snímky. Pre rýchly a jednotný prístup ku dátam používame dátový formát netCDF4, kde sú obrazové dáta uložené v troch dimenziách – obrazové riadky, stĺpce a čas [11]. V podpornej relačnej databáze sú uložené kalibračné koeficienty pre každú snímku, geometrické a iné podporné údaje. Pri načítaní dát sú využívané dodatočné geometrické korekcie zabezpečujúce korigovanie malých posunov obrazu a distorzie. Proces spracovania surových dát do harmonizovanej databázy má teda nasledovné kroky, **Obr. 2**.



Obr. 2. Diagram ilustrujúci hlavné moduly harmonizovanej databázy diaľkového prieskumu zeme. Pre každý senzor aplikačné moduly naprogramované v programovacom jazyku Python zohľadňujú ich špecifiká. Čítanie časovo-priestorového dátového bloku a časového radu meraní pre jeden bod sú už unifikované.

V takto nadizajnovanej databáze sú uložené archívne dáta pre 12 a viac rokov podľa **Tab. 1**. Nástroje umožňujú presne rovnakým procesom aj neustálu aktualizáciu databázy čím je možné zapracovávať a zohľadniť pre odvodené produkty aj najnovšie dáta.



Obr. 3. Ukážka pre vybraný bod vo viditeľnom spektre pre 11-ročné obdobie z MFG satelitov. Ide o stredomorskú oblasť takže môžeme pozorovať väčšie množstvo oblačnosti na jar a začiatkom leta. Operatívna kalibrácia zabezpečuje konverziu „digital counts“ do „true radiance“ vo fyzikálnych jednotkách $W/m^2/sr$. Vidíme, že je korigovaná znížená citlivosť senzora počas rokov a tiež je zabezpečená interkalibrácia medzi rôznymi senzormi [12, 13].

5 ZÁVER

Satelitné centrá väčšinou poskytujú archívne dáta prostredníctvom on-line dátových portálov, alebo sú distribuované cez telekomunikačné satelity. Najlepšie dátové centrá majú európsky Eumetsat a americká NOAA. Pre záchranu starých dát za účelom vytvárania klimatických databáz z teraz už 30-ročnej histórie meteorologických satelitov boli realizované viaceré medzinárodné projekty. Jedným z nich je napríklad „International Satellite Cloud Climatology Project (ISCCP) [14], kde sú z originálnych dát odvodené tzv. B1, B2 dáta so zníženým geometrickým aj časovým rozlíšením. Dáta sú interkalibrované pre celý svet na úrovni originálnych hodnôt a sú určené práve na klimatologické štúdie. Tieto dáta sú ale oklieštené o cenné metainformácie, sú tu aplikované sekundárne kalibrácie a úpravy čo znemožňuje použitie špecifických korekcií pre každý senzor. Vytvorenie vlastnej databázy z primárnych satelitných dát nám umožňuje vytvárať produkty oveľa vyššej kvality, máme detailné znalosti o každom využitom satelite, ich senzoroch a ich konštrukcii. To nám umožňuje v budúcnosti adaptovať nové senzory a dátové formáty bez viazanosti na iné medzinárodné výskumné tímy. Navyše máme spoľahlivú back-end databázu umožňujúcu:

- harmonizovaný a rýchly prístup ku satelitným dátam
- sú integrované dáta z ľubovoľných senzorov s rôznymi vlastnosťami
- je možná aktualizácia v reálnom čase

Vysoko dostupná databáza diaľkového prieskumu zeme predstavuje základný back-end systém pre solárne modely a následné poskytovanie odvodených produktov našim zákazníkom cez portál SolarGIS [15, 16].

LITERATURA

1. EUMETSAT. Earth Observation Portal, <http://archive.eumetsat.int/>, 29.11.2011
2. NOAA Comprehensive Large Array-Data Stewardship System (CLASS), <http://www.class.ngdc.noaa.gov/>, 29.11.2011
3. EUMETSAT. Meteosat First Generation, <http://www.eumetsat.int/Home/Main/Satellites/MeteosatFirstGeneration>, 29.11.2011
4. EUMETSAT. Meteosat Second Generation, <http://www.eumetsat.int/Home/Main/Satellites/MeteosatSecondGeneration>, 29.11.2011
5. University of Illinois. History of GOES, [http://ww2010.atmos.uiuc.edu/\(Gh\)/guides](http://ww2010.atmos.uiuc.edu/(Gh)/guides), 29.11.2011
6. DRL 504-02, Part 1, October 2005, Geostationary Operational Environmental Satellite (GOES) Operational Ground Equipment (OGE) Interface Specification, U.S.Department of Commerce, NOAA/NESDIS.
7. NASA, National Space Science Data Center (NSSDC) Master Catalog, <http://nssdc.gsfc.nasa.gov/nmc/>, 29.11.2011
8. CGMS 03 Coordination Group for Meteorological Satellites, 12 August 1999, LRIT/HRIT Global Specification, EUMETSAT, Am Kavalleriesand 31, P.O.Box 100555, D-64205 Darmstadt, Germany.
9. EUM FG 1, April 2000, Basic Imagery OpenMTP Format, EUMETSAT - Am Kavalleriesand 31, D-64295 Darmstadt, Germany.
10. University of Wisconsin – Madison, 1997, New AREA Format Information, Space Science and Engineering Center, University of Wisconsin – Madison, 1225 West Dayton Street, Madison, WI 53706, USA.
11. UNIDATA, NetCDF (Network Common Data Form), <http://www.unidata.ucar.edu/software/netcdf>, 29.11.2011
12. EUMETSAT, MFG Calibration, <http://www.eumetsat.int/Home/Main/DataProducts/Calibration/MFGCalibration>, 29.11.2011
13. Govaerts, Y. M., M. Clerici, N. Clerbaux (2004). "Operational Calibration of the Meteosat Radiometer VIS Band." IEEE Transactions on Geoscience and Remote Sensing 42(9): 1900-1914.
14. K.R. Knapp (2008). „Scientific data stewardship of International Satellite Cloud Climatology Project B1 global geostationary observations.“ Journal of Applied Remote Sensing, Vol.2, 023548.
15. GeoModel Solar, SolarGIS, <http://solargis.info/>, 29.11.2011
16. Šuri M., Cebecauer T., Skoczek A., 2011. SolarGIS: Solar Data And Online Applications For PV Planning And Performance Assessment. 26th European Photovoltaics Solar Energy Conference, September 2011, Hamburg, Germany.