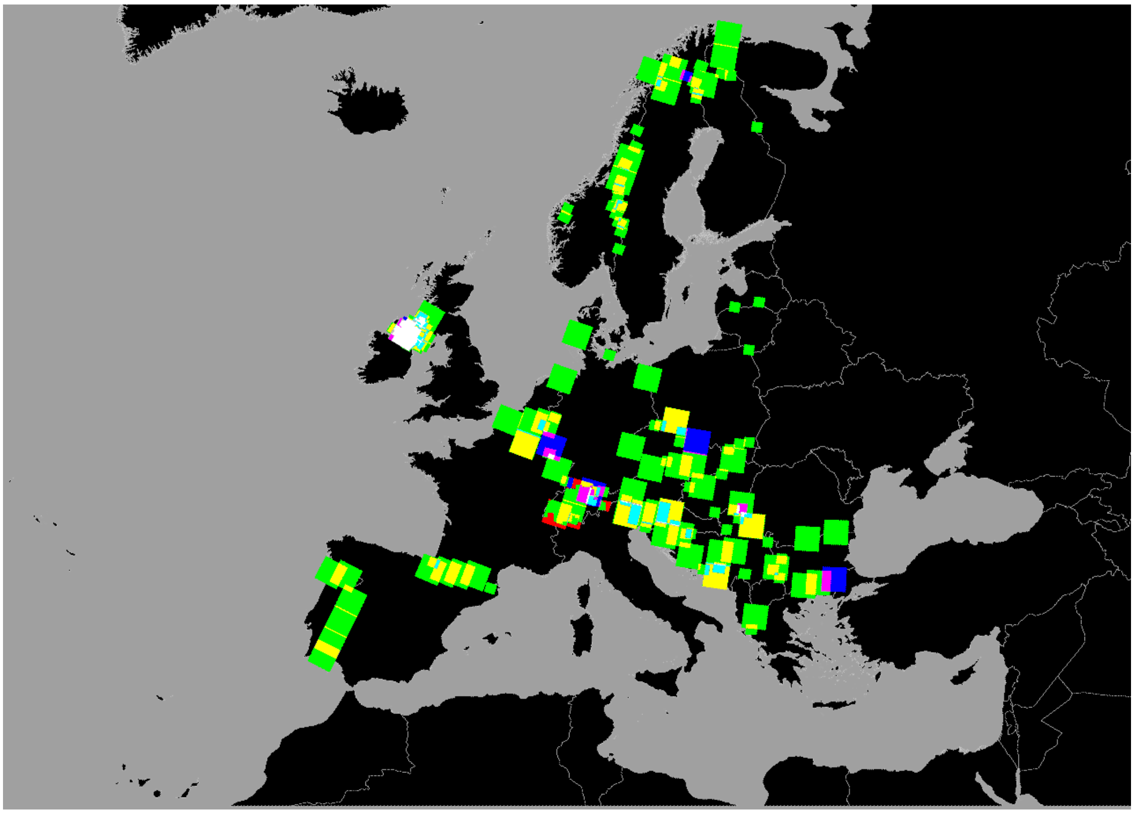
JRC Technical Notes

IMAGE-2006 Mosaic: Data Ingestion and Organisation

v.1.0

Pierre Soille and Conrad Bielski



EUR 23636 EN

EUROPEAN COMMISSION

Institute for Environment and Sustainability

The mission of the Institute for Environment and Sustainability is to provide scientific-technical support to the European Union's Policies for the protection and sustainable development of the European and global environment.

European Commission
Joint Research Centre
Institute for Environment and Sustainability
Spatial Data Infrastructures Unit

Contact information
Address: via E. Fermi 2749, I-21027 Ispra (Italy)
E-mail: Pierre.Soille@jrc.ec.europa.eu
Tel.: int+39-0332 785 068
Fax: int+39-0332 786 325

http://ies.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC 47125

EUR 23636 EN
ISBN 978-92-79-20958-1
ISSN 1831-9424
doi:10.2788/50518

Luxembourg: Publications Office of the European Union, 2011

IMAGE-2006 Mosaic: Data Ingestion and Organisation —v.1.0—

Pierre Soille and Conrad Bielski
COSIN Action
IES-Spatial Data Infrastructures Unit

September 24, 2008

Abstract

This report details how the IMAGE-2006 data was ingested and organised in view of creating the IMAGE-2006 mosaic. In particular, it details the method developed for merging the two coverages (delivered on a country basis) into a unique pan-European coverage. The concept of data and country regions of interest is introduced and a method for compositing identical scenes originating from more than one country is detailed. The resulting reference coverage contains 3,533 unique scenes out of a total of 3,699 delivered scenes.

Contents

1 Received data

The JRC received the data directly from DLR on portable hard disks on the 28th of February 2008 with a series of subsequent updates through FTP. The last update, on the 15th of April 2008, corrected for all the imagery of the Czech Republic because it did not follow the European reference grid specification [2]. The received data after this last update is referred to as VER1-0. A technical report describing procedures, methodologies, activities, and final results including statistics to document the orthoimage product generation project has been made available to JRC on the 14th of July 2008 [10].

The received data is organised on a per country basis with two coverages containing each imagery in European and national projections. The countries are referred to by the following country codes using lower case letters: al at ba be bg ch cs cy cz de dk ee es fi fr gb gr hr hu ie is it li lt lu lv mc me mt ni nl no pl pt ro se si sk tr. Note that there are 39 codes instead of 38 (actual number of participating countries) because the usual code for United Kingdom (UK) was divided into Great Britain (gb) and Northern Ireland (ni). This distinction originates from a special request from United Kingdom to process Northern Ireland in the Irish projection [10]. Note also that the code used for the Former Yugoslav Republic of Macedonia (FYROM) is mc and not the normalised ISO code MK [8]. Similarly, the code used for the republic of Serbia (cs) does not match the ISO code RS (cs corresponds to the code used for the former union of Serbia and Montenegro). Figure 1 illustrates how the received data is organised for the case of Albania (country code al). Only the European and National directories are populated with imagery data. This data is encapsulated in an archive file (zip format), one for each image of the considered country and coverage. Figure 2 lists the content of sample IRS and SPOT archive files as stored in the European sub-directories..

The `imagery.bil` file contains the orthorectified imagery in either the ETRS-LAEA projection [3] following the European grid specification [2] (`European` directory) or the national projection (`National` directory). The `geolayer.bil` refers to the geolocation file. This two band file indicates the x-y coordinates of each pixel of the non-orthorectified (raw) imagery in the target projection (either European or National). Unfortunately, it has little value at the moment because the raw imagery was not included in the `Original_IRSP6` and `Original_Spot` directories. For IRS-P6, comprehensive metadata information related to the raw imagery is stored in an ASCII file in the Super Structure Data (SSD) format [7] while for SPOT this information is stored in an XML ASCII file in the Digital Image MAP (DIMAP) format [4]. Further to image data, the Ground Control Point (GCP) image chip database [10] downloaded from DLR by ftp on the 21st of February 2008 is stored under the directory `GCP` under each country directory. This database contains 61,053 chips of size 101×101 pixels stored in `png` format. The name of each chip is in the form `Easting_Northing_EllipsoidHeight.png`.

All received data described in this section is available to the IES SDI and LMNH units at the following location: `netapp2:/geodata/EUIM/2006_DLR`. The received imagery in the ETRS-LAEA projection is summarised in table 1 for each coverage, for each sensor, and for both coverages. The number between parenthesis indicates the number of unique source images. Indeed, the imagery for the European coverage has been delivered and processed on a country basis

```
2006_DLR/VER1-0/
|---al
|-----Coverage_1
|-------European
|-------National
|-------Original_IRSP6
|-------Original_Spot
|-----Coverage_2
|-------European
|-------National
|-------Original_IRSP6
|-------Original_Spot
|-----GCP
```

Figure 1: Directory tree structure as received from DLR: example for Albania (country code al). The GCPs were received separately and were added to each country directory.

```
Directory:  2006_DLR/VER1-0/cz/Coverage_1/European/
Archive File:
IR06_LI3_ORT_10_20050922T101420_20050922T101442_DLR_10035_PREU.BIL.ZIP
  Length     Date   Time    Name
 --------    ----   ----    ----
222874960  04-15-08 10:55   imagery.bil
      150  04-15-08 10:55   imagery.blw
      120  04-15-08 10:55   imagery.hdr
407375360  04-15-08 10:57   geolayer.bil
     2411  04-15-08 10:57   metadata.xml
      672  04-15-08 10:54   calib.dat
  1412334  04-15-08 10:54   050922P6029032L0000S3_ssd.txt
 --------                   -------
631666007                   7 files

Directory:  2006_DLR/VER1-0/li/Coverage_2/European
Archive:
SP04_HRV1_X__10_20060615T102622_20060615T102631_DLR_227_PREU.BIL.ZIP
  Length     Date   Time    Name
 --------    ----   ----    ----
 33876640  10-27-07 19:03   imagery.bil
      150  10-27-07 19:03   imagery.blw
      120  10-27-07 19:03   imagery.hdr
 72000000  10-27-07 19:03   geolayer.bil
     2376  01-23-08 14:00   metadata.xml
     1596  10-27-07 19:03   calib.dat
  2048190  10-27-07 19:03   metadata.dim
 --------                   -------
107929072                   7 files
```

Figure 2: List of files contained in an archive file: examples for IRS and SPOT.

Table 1: Total number of images per sensor for first and second coverages. The number between parenthesis indicates the number of unique source images. Indeed, the image for the European coverage has been received on a country basis so that some images have been orthorectified and delivered more than once. Status as of Thu Apr 24 13:53:15 CEST 2008 (39 countries).

	1st coverage	2nd coverage	combined coverages
SPOT4	861 (836)	663 (633)	1524 (1453)
SPOT5	552 (541)	340 (335)	892 (871)
IRS-P6	667 (627)	616 (593)	1283 (1209)
total	2080 (2004)	1619 (1561)	3699 (3533)

so that a scene overlapping two or more countries may have been orthorectified and delivered more than once with different ground control points and/or reference data. It follows that duplicate scenes are not necessarily identical (in geometry and sometimes even in data footprint extent). Consequently, the data referred to as European corresponds to a union of national data made available in the ETRS-LAEA projection. There is therefore a risk that the final European mosaic will come down to a union of 38 country mosaics rather than a truly pan-European mosaic. Note that a total of 166 scenes were delivered two or three times. The former case occurred 148 times and the later 9 times (i.e., $166 = 148 + 9 \times 2$).

All duplicate scenes were processed at a later stage in such a way that the final number of 3,533 unique scenes was reached (see Sec. 5). This final set of unique images is referred to as the *combined* or *reference coverage*. Given the country based processing, the quality of the orthorectification of an image is only guaranteed within the country this image belongs to[1]. This feature is taken into account when determining the domain of reliability of the data masks of each image, see Sec. 5. Finally, table 2 details the number of images received per country, per sensor, and per coverage. The master list with scene names per country and coverage was made available to the JRC on the 12th of August 2008. This list is identical to the list reconstructed from the received data (apart from the 21 ni scenes that were not included in the master list).

2 Naming convention

The heterogeneity of the input imagery motivated us to develop a naming convention suitable for any type of sensor [18]. Therefore, it should not rely on sensor specific organisation such as the path/row numbers of Landsat and IRS satellites. A generic solution is achieved by including the x-y coordinates of the centre of the scene in the name of the file corresponding to that scene. The coordinates are given in the ETRS-LAEA projection. Files are themselves organised according to the European reference grid [2]. Figure 3 shows a graphical representation of the levels 18 and 19 of the European grid. Given the spatial resolution of the imagery delivered in the LAEA projection (25m), we have opted for

[1]A detailed assessment of the relative geometric accuracy between all pairs of overlapping scenes is presented in [16].

Table 2: Total number scenes received per country (country codes used in data delivered by DLR [10]), per coverage, and per sensor. Corresponding to the data sent by DLR on the 26th of February and received at JRC on the 28th of February 2008.

	coverage 1				coverage 2				combined coverages			
	SP4	SP5	IL3	ALL	SP4	SP5	IL3	ALL	SP4	SP5	IL3	ALL
al	8	1	5	**14**	2	1	4	**7**	10	2	9	**21**
at	15	5	12	**32**	11	4	15	**30**	26	9	27	**62**
ba	3	5	9	**17**	3	8	6	**17**	6	13	15	**34**
be	3	6	5	**14**	4	5	3	**12**	7	11	8	**26**
bg	24	11	12	**47**	11	8	12	**31**	35	19	24	**78**
ch	11	6	5	**22**	8	0	7	**15**	19	6	12	**37**
cs	21	11	8	**40**	9	2	12	**23**	30	13	20	**63**
cy	0	1	2	**3**	0	0	2	**2**	0	1	4	**5**
cz	10	8	8	**26**	7	3	13	**23**	17	11	21	**49**
de	48	52	44	**144**	15	7	41	**63**	63	59	85	**207**
dk	18	14	4	**36**	11	6	8	**25**	29	20	12	**61**
ee	12	3	7	**22**	0	1	8	**9**	12	4	15	**31**
es	21	23	59	**103**	15	26	65	**106**	36	49	124	**209**
fi	34	6	47	**87**	37	3	36	**76**	71	9	83	**163**
fr	95	50	47	**192**	23	10	62	**95**	118	60	109	**287**
gb	85	23	23	**131**	99	56	15	**170**	184	79	38	**301**
gr	45	30	21	**96**	12	13	34	**59**	57	43	55	**155**
hr	2	10	10	**22**	4	6	10	**20**	6	16	20	**42**
hu	20	16	5	**41**	8	1	10	**19**	28	17	15	**60**
ie	16	13	10	**39**	25	8	6	**39**	41	21	16	**78**
is	59	17	0	**76**	0	0	0	**0**	59	17	0	**76**
it	23	34	37	**94**	48	44	41	**133**	71	78	78	**227**
li	0	0	1	**1**	1	0	0	**1**	1	0	1	**2**
lt	9	17	4	**30**	11	3	7	**21**	20	20	11	**51**
lu	0	2	1	**3**	3	0	0	**3**	3	2	1	**6**
lv	18	2	6	**26**	3	1	8	**12**	21	3	14	**38**
mc	3	0	6	**9**	3	1	3	**7**	6	1	9	**16**
me	3	0	3	**6**	3	1	2	**6**	6	1	5	**12**
mt	0	0	1	**1**	0	1	0	**1**	0	1	1	**2**
ni	3	0	3	**6**	12	2	1	**15**	15	2	4	**21**
nl	11	13	5	**29**	15	8	2	**25**	26	21	7	**54**
no	62	9	57	**128**	129	32	6	**167**	191	41	63	**295**
pl	52	10	29	**91**	14	5	31	**50**	66	15	60	**141**
pt	10	5	12	**27**	19	5	10	**34**	29	10	22	**61**
ro	22	22	24	**68**	31	11	21	**63**	53	33	45	**131**
se	49	66	42	**157**	27	15	37	**79**	76	81	79	**236**
si	6	3	2	**11**	4	3	3	**10**	10	6	5	**21**
sk	12	5	5	**22**	2	3	6	**11**	14	8	11	**33**
tr	28	53	86	**167**	34	37	69	**140**	62	90	155	**307**
ALL	**861**	**552**	**667**	**2080**	**663**	**340**	**616**	**1619**	**1524**	**892**	**1283**	**3699**

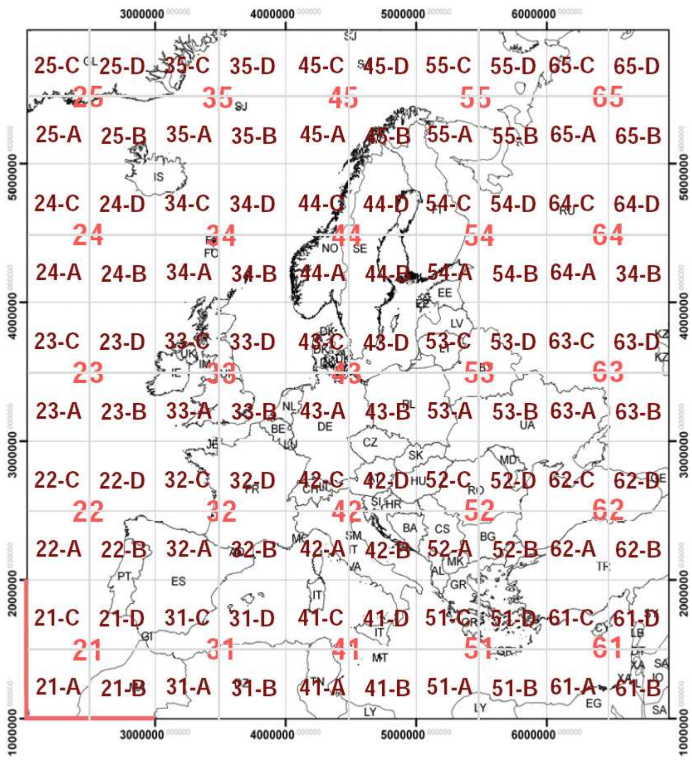


Figure 3: European grid level 18 and 19: grid space 500km × 500km (level 18) and 1000km × 1000km (level 19). Source: [18].

tiles at the grid level 18. That is, each file is stored in a sub-directory whose tile name is defined by the coordinates of the centre of the corresponding scene. For example, a file whose centre coordinate belongs to the tile 42-D, is stored in the directory named `42-D`. The full file names are built by concatenating 4 underscore separated subfields: `YYYYMMDD-hhmm_sss_tt-iiii_xxxxxyyyyy-qq.tif`. The meaning of each subfield is described hereafter [18]:

- `YYYYMMDD-hhmm`: reference date and time;
- `sss`: sensor acronym (e.g., SP4 for SPOT4, SP5 for SPOT5, IL3 for IRS LISS 3, and DMC for Disaster Monitoring Constellation);
- `tt-iii`: `tt` for data type composed of two letters, the first being either I (individual) or C (composite) and the second being either P (physical parameter, e.g. digital number or top of atmosphere reflectance), M (mask, e.g. segment labels or classified image), or X (mixture for P and M combined into a single multiband file) and `iii` for identifier with no restriction on length. For example, we use the identifier IP-B1XX for the first band of a SPOT or IRS image delivered for the country with country code XX. This identifier is unique even in case the same scene was delivered for more than one country. Once identical imagery originating from more than one country or coverage are merged to create the reference coverage (see Sec. 6), we use the identifier IP-B10C for the first band of a SPOT or

IRS image whatever the name of the country or countries it was delivered for (this information can be retrieved from the header file described in Sec. 4).

- `xxxxxyyyyy-qq`: reference pixel cell code indicating the median of the area filled with data values (see details in Sec. 5.1). For example, the reference pixel cell code of a scene with centre in Sicily will always be in the form `45xxx15yyy-qq` because Sicily lies fully in the level 18 tile `41-D`, see Fig. 3.

3 Storage format

For storing ingested imagery data, we have chosen the GeoTIFF [12] format given its flexibility and high portability. This format represents an effort by over 160 different remote sensing, GIS, cartographic, and surveying related companies and organisations to establish a TIFF [1] based interchange format for georeferenced raster imagery. TIFF allows for embedded lossless compression. We have opted for the LZW compression [20] with horizontal differencing. Rather than relying on specific coding systems such as the EPSG codes [5], all projection parameters are given in full. Figure 4 lists the parameters as extracted from a sample ETRS-LAEA file.

4 Metadata

While the GeoTIFF files contain all necessary information for successfully reading an image including all information regarding its associated coordinate reference system and position within this system, the GeoTIFF format does not yet include standardised tags for storing additional metadata such as those required for transforming the provided digital numbers into reflectance values. For this reason, it was decided to accompany each GeoTIFF file with a header file following the ENVI specifications (in ASCII format). Apart form the ENVI required fields, additional useful metadata was included. This additional information begins with a semi-colon (indicating a comment for ENVI), then the name of the field and its value. The content of a typical header file is shown in Fig. 5 for a SPOT image and Fig. 6 for an IRS image. Apart from the path and row fields only available for IRS-P6 images, these header files contain the same fields. The at-sensor solar exoatmospheric irradiance values were extracted from the delivered metadata file for the SPOT sensors[2] and from [11, 6] for the IRS-P6 sensor.

5 Data/no data masking

The delivered data does not contain information that could lead directly to data/no data masks for each delivered image. To address this lack of information, an ad hoc (indirect) procedure was developed. This procedure consists of two main steps. The first solves the data/no data masking per se. The second step addresses the fact that scenes were processed on a country basis (so

[2] see also [9] and `http://www.spotimage.fr/automne_modules_files/standard/public/p229_fileLINKEDFILE_new_lumina.PDF`.

```
Geotiff_Information:
   Version: 1
   Key_Revision: 1.0
   Tagged_Information:
      ModelTiepointTag (2,3):
         0                0                0
         916475           2825275          0
      ModelPixelScaleTag (1,3):
         25               25               0
      End_Of_Tags.
   Keyed_Information:
      GTModelTypeGeoKey (Short,1): ModelTypeProjected
      GTRasterTypeGeoKey (Short,1): RasterPixelIsArea
      GTCitationGeoKey (Ascii,9): "JRC-EGCS"
      GeographicTypeGeoKey (Short,1): GCS_EUREF89
      GeogCitationGeoKey (Ascii,7): "ETRS89"
      GeogGeodeticDatumGeoKey (Short,1): Datum_European_Reference_System_1989
      GeogPrimeMeridianGeoKey (Double,1): 3.16640047e-269
      GeogLinearUnitsGeoKey (Short,1): Linear_Meter
      GeogAngularUnitsGeoKey (Short,1): Angular_Degree
      GeogEllipsoidGeoKey (Short,1): Ellipse_GRS_1980
      GeogSemiMajorAxisGeoKey (Double,1): 6378137
      GeogInvFlatteningGeoKey (Double,1): 298.257222
      GeogPrimeMeridianLongGeoKey (Double,1): 0
      ProjectedCSTypeGeoKey (Short,1): Unknown-3035
      PCSCitationGeoKey (Ascii,19): "ETRS89 / ETRS-LAEA"
      ProjectionGeoKey (Short,1): User-Defined
      ProjCoordTransGeoKey (Short,1): CT_LambertAzimEqualArea
      ProjLinearUnitsGeoKey (Short,1): Linear_Meter
      ProjFalseEastingGeoKey (Double,1): 4321000
      ProjFalseNorthingGeoKey (Double,1): 3210000
      ProjCenterLongGeoKey (Double,1): 10
      ProjCenterLatGeoKey (Double,1): 52
      End_Of_Keys.
   End_Of_Geotiff.

PCS = 3035 (ETRS89 / ETRS-LAEA)
Projection = 19986 (Europe Equal Area 2001)
Projection Method: CT_LambertAzimEqualArea
   ProjCenterLatGeoKey: 52.000000 ( 52d 0' 0.00"N)
   ProjCenterLongGeoKey: 10.000000 ( 10d 0' 0.00"E)
   ProjFalseEastingGeoKey: 4321000.000000 m
   ProjFalseNorthingGeoKey: 3210000.000000 m
GCS: 4258/ETRS89
Datum: 6258/European Terrestrial Reference System 1989
Ellipsoid: 7019/GRS 1980 (6378137.00,6356752.31)
Prime Meridian: 8901/Greenwich (0.000000/  0d 0' 0.00"E)
Projection Linear Units: 9001/metre (1.000000m)

Corner Coordinates:
Upper Left    (  916475.000, 2825275.000)  ( 31d52'24.43"W, 39d42' 3.18"N)
Lower Left    (  916475.000, 2739625.000)  ( 31d23'37.09"W, 39d 3'24.40"N)
Upper Right   ( 1000350.000, 2825275.000)  ( 31d 0' 6.36"W, 40d 6'12.09"N)
Lower Right   ( 1000350.000, 2739625.000)  ( 30d31'35.49"W, 39d27'16.09"N)
Center        (  958412.500, 2782450.000)  ( 31d11'53.78"W, 39d34'49.29"N)
```

Figure 4: GeoTIFF header information on a sample file 20070823-1303_SP4_IP-B10C_0958427824-CA.tif indicating all available geoinformation (output of the `listgeo` command).

```
ENVI
description = { Image2006 - ingested on 2008-05-18 based on ingestIM2K6 v1.2}
samples = 3355
lines = 3426
bands = 1
header offset = 0
file type = TIFF
data type = 1
interleave = bsq
sensor type = SPOT
byte order = 0
read procedures = {idl_tiff_read_spatial, idl_tiff_read_spectral}
map info = {ETRS-LAEA, 1.0000, 1.0000,916475.0, 2825275.0, 25.0, 25.0, ETRS89, units=Meters}
projection info = {11, 6378137.0, 6356752.3, 52.000000, 10.000000, 4321000.0, 3210000.0,
                   ETRS89, ETRS-LAEA, units=Meters}
band names = {SP4 B1GRN}
data ignore value = 0
default stretch = 2.0% linear
wavelength units = Micrometers
wavelength = {0.545}
bbl = {1}
default bands = {1}
data gain values = { 0.3344 }
data offset values = { 0 }
;sunElevation = 58.9
;sunAzimuth = 148.4
;acquisitionDate = 20070823
;acquisitionTime = 1303
;resampling = { CC }
;nominalAltitude = 830462.99673
;solarIrradianceValue = 1851
;imageNoDataFraction = 0.5
;countryOrigin = pt
;originalMetadata = 60260001_1A_DVD.ZIP
;parentFile = SP04_HRV2_X__1O_20070823T130328_20070823T130337_DLR_115_PREUpt
;dataSet = Image2006Coverage1LAEA
```

Figure 5: SPOT header file (`20070823-1303_SP4_IP-B10C_0958427824-CA.hdr`).

```
ENVI
description = { Image2006 - ingested on 2008-05-16 based on ingestIM2K6 v1.0}
samples = 7728
lines = 7746
bands = 1
header offset = 0
file type = TIFF
data type = 1
interleave = bsq
sensor type = IRS LISSIII
byte order = 0
read procedures = {idl_tiff_read_spatial, idl_tiff_read_spectral}
map info = {ETRS-LAEA, 1.0000, 1.0000,3537850.0, 2584450.0, 25.0, 25.0, ETRS89, units=Meters}
projection info = {11, 6378137.0, 6356752.3, 52.000000, 10.000000, 4321000.0, 3210000.0,
                   ETRS89, ETRS-LAEA, units=Meters}
band names = {IL3 B1GRN}
data ignore value = 0
default stretch = 2.0% linear
wavelength units = Micrometers
wavelength = {0.555}
bbl = {1}
default bands = {1}
data gain values = { 0.724353 }
data offset values = { 0 }
;maxgray = 255
;sunElevation = 54
;sunAzimuth = 156.5
;path = 19
;row = 37
;acquisitionDate = 20070419
;acquisitionTime = 1058
;nominalPixelSpacing = 23.500
;nominalLineSpacing = 23.500
;resampling = { CC }
;nominalAltitude = 817000.0000000
;solarIrradianceValue = 1849.50
;imageNoDataFraction = 0.46
;countryOrigin = fr
;originalMetadata = 070419P6019037L0000S3.ZIP
;parentFile = IR06_LI3_ORT_10_20070419T105837_20070419T105859_DLR_18191_PREUfr
;dataSet = Image2006Coverage2LAEA
```

Figure 6: IRS-LISSIII header file (20070419-1058_IL3_IP-B10C_3634924873-CB.hdr).

that there is unfortunately no guarantee that a scene delivered for a country is reliable outside the corresponding country's territory).

5.1 Where are data points?

Since no specific digital number (DN) was reserved for no data values and no data masks were not available, the generation of a mask indicating where actual data measurements are available can only be estimated. This is achieved with the following procedure:

- threshold each band for values ≥ 1 and sum up the resulting bands;
- threshold the resulting image for values ≥ 2. That is, a pixel is deemed to be a data point if this pixel has a DN value greater than 0 for at least two bands (theoretically, a non zero value in one band would be sufficient);
- fill the holes of the resulting image;
- erode the resulting image by a 3×3 square since the interpolated values along the border of this mask are corrupted by boundary effects.

The resulting data masks (also called Regions Of Interest or ROI) are named using the following data type and identifier field:

- `IM-DRXX` for delivered data where XX refers to the country code the image was delivered for;
- `IM-DROI` for the generated reference coverage.

All orthorectified images were delivered with data within their full original footprint except for Switzerland where large parts falling outside this territory were suppressed leading to non convex data footprints. The discrepancy between the Swiss data footprints and those obtained for all other countries is illustrated in Fig. 7 using Slovenia as example for comparison purposes.

5.2 Which data points are reliable?

Reliable data can be defined as data that was processed in such a way that the errors against the target precision are within predefined limits. Due to the fact that images we processed on a country basis without ensuring that the accuracy requirements were met outside the territory of the processed country, an image originating from a given country is only geometrically reliable within the boundaries of this country.

The domain of reliability of each received image is defined by the intersection of the full footprint (`IM-DRXX` described above) with the territory of the country the image originates from. However, for generating a European mosaic we need to allow overlap between countries because otherwise the European mosaic would come down to a simple union of country mosaics. The data has therefore been deemed as reliable when it falls within the country boundary enlarged by a buffer of 5km (400 pixels at a resolution of 25m). In addition, we have also included the sea and all parts falling outside the 39 participating countries (keeping in mind that the quality of the orthorectification is not secured in

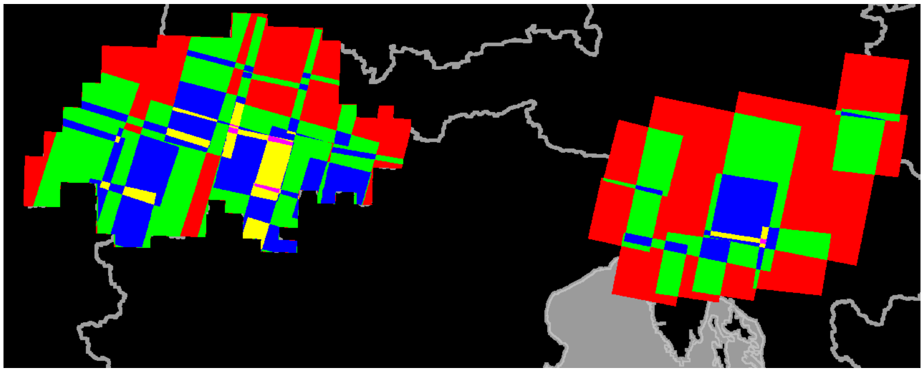


Figure 7: Point-wise maximum mosaic of the Swiss and Slovenian images for all images delivered for these countries in coverage 1. Colours are used for coding the degree of overlap using the order red, green, blue, yellow, magenta (i.e., no overlap in red and 5 images overlapping in magenta). Observe the non-convex data footprints in the case of Switzerland.

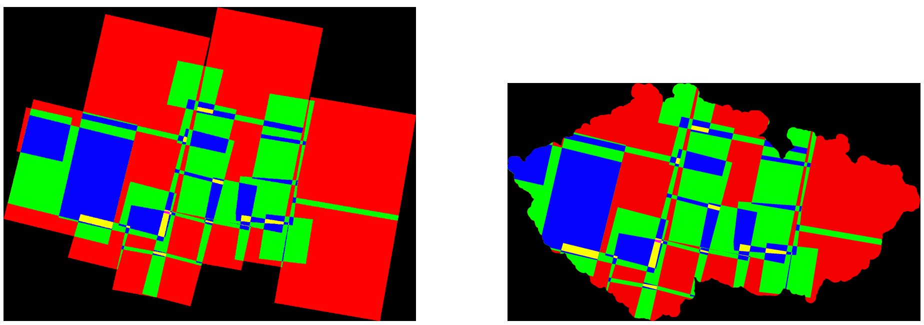


Figure 8: Data versus country regions of interest in the case of the Czech Republic. Left: `IM-DRCZ` Right: `IM-CRCZ`.

the non-participating countries). The resulting domains of reliability for the delivered data are referred to as `IM-CRXX`:

$$\texttt{IM-CRXX} = \texttt{IM-DRXX} \bigcap \left(\delta_B(\texttt{XX}) \bigcup \text{sea} \bigcup \overline{\cup_{\texttt{YY=AL}}^{\texttt{YY=TR}} \texttt{YY}} \right),$$

where `XX` denotes the country code the image was delivered for, δ the dilation operator, B a disk of radius equal to 5km, and with the list of participating countries starting with the code `AL` (for Albania) and terminates with the code `TR` (for Turkey). That is, $\overline{\cup_{\texttt{YY=AL}}^{\texttt{YY=TR}} \texttt{YY}}$ refers to the territory *not* covered by the union of the territories of the participating countries. Figure 8 illustrates the difference between data and country regions of interest in the case of the Czech Republic.

All ROI computations involving country boundaries were performed using National Territory Units at level 0 (Gisco version 9) rasterised at a spatial resolution of 25m in the European reference grid. The domains of reliability for the images of the reference coverage are referred to as `IM-CROI`. When they correspond to images that were delivered for only one country, they are identical to their corresponding `IM-CRXX`. A more general definition, holding for images that were delivered for more than one country, is to define `IM-CROI`s as the union of the domains of reliability of all input delivered images corresponding to the considered scene:

$$\texttt{IM-CROI} = \bigcup_{\texttt{YY}} \texttt{IM-CRYY},$$

where the union spans over the set of countries for which the considered image was delivered (i.e., either 1, 2 or 3 countrie(s) for either unique, duplicate or triplicate images). An equivalent formula is used for creating the reference DROIs from the coverage 1 and 2 DRXXs. An example is detailed in the next section when describing the methodology developed for merging duplicate images when creating the reference coverage.

6 Combining duplicate images

The proposed methodology for combining duplicate images is first described. It is then applied to all 3,699 delivered images so as to create a reference coverage containing 3,533 unique scenes.

6.1 Methodology

Because images were processed and delivered on a country basis, multiple delivery of a number of scenes has occurred (up to 3 deliveries of the same scene). Differences in processing have led to differences between duplicates so that arbitrary selection of an image among the duplicates is not a viable solution. For example, Figure 9 shows on the top row a triplicate SPOT4 scene included in the Swedish, Finnish, and Norwegian coverages (all in coverage 2). The date and time strings of the input 3 scenes is as follows: `20060915-1051`, the sensor name is `SP4`, and the centre pixel value is `4834251055-BC`. The rightmost image of the top row shows the territory associated with each country (light grey for Sweden, black for Finland, and dark grey for Norway). The reliable data area associated with each scene are shown in the middle row (i.e., buffered country zones). They corresponding to `IM-CRSE`, `IM-CRFI`, and `IM-CRNO` respectively. Note that the delivered scenes were processed separately leading to discrepancies as highlighted by the zoom section (bottom row).

Therefore, original scenes that were delivered for more than one country need to be composed so as to create unique scenes (this process could be referred to as *image conflation*). A simple cut along the country boundaries is not necessarily a good solution because the cutline may be visible in the output image. A better solution is to find a cut line following image structures near the country border (near in the sense that it should lie within the buffer considered when creating the country ROIs). This is achieved by adapting the morphological compositing procedure described in [15] (the resulting methodology could be referred to as *morphological image conflation*). An example is illustrated in Fig. 10 and corresponds to the compositing of the triplicate imagery shown in Fig. 9. More precisely, regions where the country ROIs do not overlap are considered as seeds for each respective image. In our example, there are 3 seeds displayed with fundamental colours on the left of Fig. 10. A region growing procedure based

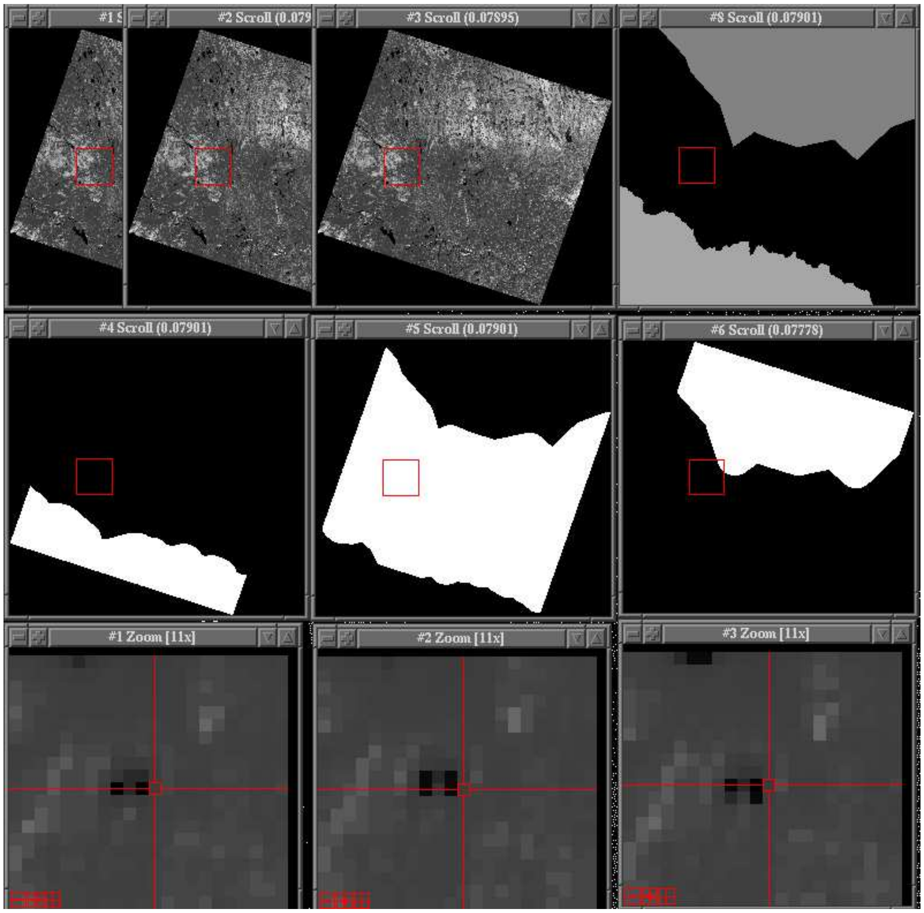


Figure 9: Example of a SPOT4 scene delivered for 3 countries (Sweden, Finland, and Norway) with the corresponding NUTS regions (top row), the resulting reliable data areas or `CRXX`s (national territory `XX` plus buffer of 5km away from the given boundary) (middle row), and a zoom highlighting the discrepancy between the 3 delivered images (bottom row).

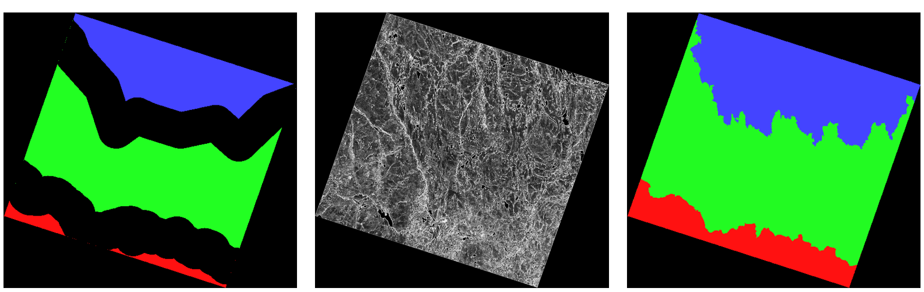


Figure 10: Composition of the triplicate scene shown in Fig. 9. Left: marker set (red for Sweden, green for Finland, and blue for Norway). Middle: pointwise minimum of the magnitude of the gradient of all three input scenes. Right: resulting decision rule obtained by propagating the markers, the propagation being driven by the gradient image.

on the watershed transformation [17, 19] is then applied to expand these seeds until the whole data ROI is covered, see right image of Fig. 10. The growth is controlled by the gradient image shown on the middle of Fig. 10. This gradient image is calculated by computing the point-wise minimum of the morphological gradient [13] of the 3rd band of all 3 input scenes. It ensures that resulting seam lines follow salient image structures. The grown seeds provide decision regions indicating which image should be used when composing the 3 input images. For example, the values of the Swedish images are selected in the red region. Further details on the composition procedure are given in [14]. The domain of reliability of the merged scene stored in the reference coverage, i.e., `IM-CROI`, is equal to the union of the 3 country based domains of reliability shown in the middle row of Fig. 9. In this example, it comes down to the full data ROI of the scene.

Finally, it should be emphasised that the proposed procedure is *ad hoc* because it does not actually fix problems resulting from orthorectification performed on a country basis with different reference data and ground control points.

6.2 Results

Tables 3 and 4 list all duplicate and triplicate scenes with their countries and coverages of origin.

Table 3: The 148 duplicate scenes among all 3,699 delivered images. In case the same centre value is found, it is given only once. Similarly, if the images originate from the same cover, the coverage number is given only once.

date-time	sensor	centre(s)	country codes	cove-rage(s)
20050401-1043	IL3	4091728422-AA	FR_DE	2
20050404-0941	IL3	5208525144-CC	CS_RO	2
20050405-0921	IL3	5593021862-AA	GR_BG	2
20050422-1005	IL3	4738725924-DA	SI_AT	2
20050422-1005	IL3	4790528576-AA_4790428576-BB	AT_CZ	2
20050422-1006	IL3	4712224600-CC	HR_SI	2
20050502-0957	IL3	4854823406-BB	BA_HR	2
20050503-0937	IL3	5228919902-DB	GR_AL	2
20050528-1054	IL3	3905729833-CB	BE_FR	2
20050528-1054	IL3	3952631099-BB	NL_BE	1_2
20050623-1012	IL3	4635728464-DB	AT_CZ	1_2
20050630-1024	SP5	4445144528-CB_4445144528-DA	SE_NO	1
20050701-1040	IL3	4718050208-AB_4718050207-CB	NO_SE	1
20050701-1040	IL3	4756451483-BA_4756451483-AB	NO_FI	1
20050702-1019	IL3	4936050599-DC_4936050600-AB	SE_FI	1
20050807-1206	SP4	3161436445-CC	GB_IE	1
20050830-1054	SP5	4017130899-BB	NL_BE	2
20050902-1021	SP4	4891349775-AB_4891349775-BA	SE_FI	2
20050909-1058	SP5	4459142083-DC	NO_SE	2
20051013-1147	SP5	3283635852-AB	GB_NI	2
20051020-1142	SP4	3328835513-CC	GB_NI	2
20051020-1142	SP4	3355935985-CB	NI_GB	2
20051026-0948	SP4	5065924928-CB	HR_CS	2
20051030-1150	SP4	3257136916-CB	NI_GB	2
20051030-1150	SP4	3265536307-BB	GB_NI	2
20051030-1150	SP4	3287636667-CB	GB_NI	2
20051107-1057	IL3	3533622371-BB	ES_FR	2
20051109-1014	IL3	4494225799-DC	IT_AT	2

date-time	sensor	centre(s)	country codes	cove-rage(s)
20051109-1014	IL3	4525727112-AB	DE_AT	2
20051109-1157	SP4	3334536562-BB	GB_NI	2
20051109-1158	SP4	3304536099-AA	GB_NI	2
20060121-1153	SP4	3233935553-CB	NI_GB	2
20060221-1156	SP4	3217536303-DB	GB_NI	2
20060221-1156	SP4	3247436763-DA	NI_GB	2
20060325-0940	IL3	5149026378-AB_5149026378-AA	RO_HU	2
20060327-0925	SP4	5208219158-DC	GR_AL	2
20060506-1004	IL3	4771931167-CB	PL_CZ	2
20060510-1157	SP4	3308937149-CC	NI_GB	1
20060511-1140	IL3	3286136008-CC	NI_GB	1
20060511-1140	IL3	3354237180-DC	GB_NI	1
20060605-1123	IL3	2909320826-BC	ES_PT	2
20060605-1124	IL3	2753917659-BB	PT_ES	2
20060605-1124	IL3	2791218418-CA	ES_PT	2
20060605-1124	IL3	2850219621-CB	PT_ES	2
20060606-1102	IL3	3453422488-BB	ES_FR	1_2
20060611-1109	SP5	4433543061-AB	SE_NO	1
20060611-1109	SP5	4455843558-BB	SE_NO	1
20060612-1119	SP4	4483345985-BB_4483345985-DB	SE_NO	2
20060612-1119	SP4	4542646971-AB_4542646971-AA	SE_NO	1
20060613-1014	IL3	4499625800-DB	IT_AT	1
20060615-1026	SP4	4288627050-BA_4305927115-CC	CH_AT	1
20060620-0929	IL3	5349522733-DB	CS_BG	1
20060622-0951	SP4	4832324676-CB	BA_HR	1
20060627-0955	SP4	5107726059-CA	CS_RO	1
20060627-1134	SP4	3318436830-DC	GB_NI	1
20060701-0946	SP5	5193235272-CB	LT_PL	1
20060702-0956	SP4	5252738022-CB	LT_LV	1
20060703-0939	SP4	5194429953-DA	SK_PL	1
20060706-0952	SP5	4861024702-DB	BA_HR	1_2
20060706-1021	SP4	4684730939-DB	DE_CZ	1
20060708-0953	IL3	4926727402-CC	SK_HU	1
20060710-1052	IL3	3908829832-CB	FR_BE	1
20060711-1031	IL3	4211327084-BB_4215226885-BC	DE_CH	1_2
20060711-1032	IL3	4165025940-DC_4172525791-DB	CH_IT	1
20060712-0937	SP5	5134725559-DC_5134725560-BA	RO_CS	1
20060713-1057	SP5	4079531391-BC	DE_NL	1
20060715-1019	SP5	4148927673-AA	DE_FR	1
20060715-1046	IL3	4120033598-AA	NL_DE	1
20060715-1046	IL3	4209036146-DC	DE_DK	1_2
20060715-1047	IL3	4030031051-CA	BE_DE	1
20060715-1047	SP4	4392534980-CC	DE_DK	1_2
20060717-1005	IL3	4770431165-DB	PL_CZ	1
20060717-1006	IL3	4660925873-CB	SI_AT	1
20060717-1011	SP4	4352827116-CA	DE_AT	1
20060717-1150	SP4	3250335691-BC	IE_GB	1
20060718-0944	IL3	5102928939-CA	SK_HU	1
20060718-0946	IL3	5005822248-DB	HR_ME	1
20060718-0950	SP4	5080329542-AC	SK_PL	1
20060719-0932	SP4	5302424194-BB	CS_BG	1_2
20060722-0944	SP5	4997525920-BB	CS_HU	1
20060722-1002	IL3	4742025924-CC	SI_AT	1
20060726-1017	IL3	4612333643-DC	DE_PL	1
20060728-0958	SP4	5139129863-CA	PL_SK	1
20060729-1050	IL3	4635651323-AA_4635651323-AB	SE_NO	1
20060803-0912	IL3	5693724759-BB	RO_BG	1
20060803-1047	IL3	4503846313-CB	NO_SE	1
20060803-1048	IL3	4459645046-CB_4459645046-CC	NO_SE	1
20060804-1031	SP5	4472942583-DA_4472942583-CB	NO_SE	1
20060806-1132	IL3	2795422440-BB	ES_PT	1
20060811-0957	SP5	5111837742-AB	LT_LV	1_2
20060811-1128	IL3	2881022233-CB	PT_ES	1
20060816-0942	IL3	5110923690-BB	CS_ME	1
20060817-0921	IL3	5528824395-CC	BG_RO	1

date-time	sensor	centre(s)	country codes	cove-rage(s)
20060817-1014	SP4	4722428854-AA	AT_CZ	1
20060818-0956	SP4	5170422365-BB	CS_ME	1
20060819-0937	SP4	5369322130-BB	BG_MC	1_2
20060819-0937	SP4	5382223243-CB	CS_BG	1
20060819-1018	IL3	4516829710-BB	DE_CZ	1
20060820-1051	SP4	4771751853-DA_4771751853-DB	NO_FI	1_2
20060824-1008	IL3	5050752143-AC_5050652143-BB	NO_FI	1_2
20060824-1008	IL3	5073253454-CA	NO_FI	1_2
20060831-1042	SP4	4463943600-AB	NO_SE	2
20060902-0937	SP5	4934022621-CB_4934022621-DC	HR_BA	1
20060904-0946	IL3	5027523587-AB	CS_BA	1
20060904-0947	IL3	5007122251-AB	BA_HR	1_2
20060904-1040	SP5	3710521540-BC	ES_FR	1
20060905-1107	IL3	3366822610-DB	FR_ES	1_2
20060906-1000	SP5	4600825055-DB	SI_HR	1_2
20060906-1030	SP4	4047529304-AA	FR_LU	1_2
20060907-1010	SP4	4796129973-DB	PL_CZ	1
20060907-1010	SP4	4807430513-BB	PL_CZ	1
20060910-1023	SP5	4488426608-CC	IT_AT	2
20060910-1100	IL3	3809431227-DC	FR_BE	1
20060913-0958	IL3	4872428656-DB	CZ_SK	1
20060913-0959	IL3	4799424669-DB	HR_BA	1
20060914-0938	IL3	5211925154-DC	RO_CS	1
20060914-1111	SP4	4551247921-BC	NO_SE	1
20060915-1051	SP4	4689950506-CA_4689950506-CC	NO_SE	1_2
20060915-1053	SP4	4456445015-DB	SE_NO	2
20060915-1054	SP4	4446341087-AC_4446341087-AB	NO_SE	1_2
20060920-1103	SP4	3385423034-AA	ES_FR	1_2
20060922-1022	SP4	4653330917-CA	PL_CZ	1_2
20060923-1003	SP4	4851227368-AA	HU_AT	1_2
20060928-1007	SP4	5038328114-BB	SK_HU	1_2
20061004-0922	IL3	5509721695-DA	BG_GR	2
20061008-1015	SP4	4352026275-CB_4365826291-DB	CH_IT	2
20061014-1054	IL3	3624422263-AB	ES_FR	2
20061016-1011	IL3	4581025833-BA	AT_IT	2
20061016-1057	SP4	4407343541-CA_4407343541-AC	SE_NO	2
20061019-1004	SP4	5023322718-AC	BA_ME	2
20061020-0945	SP4	5331922611-AB	BG_CS	2
20061023-1206	SP4	3287435601-AC	GB_NI	2
20061103-1037	IL3	4092725816-BC_4102426004-AA	FR_CH	2
20061103-1154	SP4	3153036465-BA	GB_IE	2
20061107-1039	SP4	4175625658-BB_4186425536-AB	CH_IT	2
20070125-1158	SP4	3183935914-BA	NI_GB	2
20070203-1153	SP5	3318737125-CB	NI_GB	2
20070501-1146	IL3	3216236173-BA	NI_GB	2
20070701-1103	SP5	4150943404-CC	NO_NO	1_2
20070704-1034	SP4	5036651191-DB	FI_FI	1_2
20070806-1000	SP4	5239748109-CC	FI_FI	1_2
20070807-0940	SP4	5088551065-AB	FI_FI	1_2
20070808-1029	SP5	4898950040-BB	FI_FI	1_2
20070822-1134	SP4	4137742927-BA	NO_NO	1_2
20070929-1100	SP4	4919751637-CB	NO_FI	2
20070929-1101	SP4	4880950684-CB_4880950684-CA	FI_SE	2
20071006-1028	SP4	4441044039-AC	NO_SE	2
20071009-1110	SP4	4450242572-BA_4450242572-BC	NO_SE	2

A graphical representation of the footprints of all duplicate imagery is shown in Fig 11. In this image, the binary footprints of all 323 duplicate scenes have been added ($323 = 148 \times 2 + 9 \times 3$). This enables the use of colours for coding the degree of overlap: red for 1 (no overlap), green for 2 (duplicates), blue for 3 (triplicates), yellow for 4 (pair of overlapping duplicates), etc. Note that the red regions appear only for scenes delivered jointly for Switzerland and a neighbour country. This is due to the non rectangular footprints of the scenes delivered for Switzerland (see also Fig. 7). Also note the two SPOT footprints on the west coast of Norway corresponding to the same scenes delivered for both coverages (no intersection with a country boundary). More precisely, six scenes

Table 4: The 9 triplicate scenes among all 3,699 delivered images. The centre pixel identifier of the third scene is in all cases identical to that of the second scene.

f

date-time	sensor	centre(s)	country codes	cove-rage(s)
20050907-1028	IL3	4288127076-AA_4288127076-AA_id.	LI_DE_AT	1_1_2
20060404-1149	SP4	3225036819-CB_3225036819-CB_id.	GB_NI_IE	1_1_1
20060523-0912	IL3	5682322053-BB_5682322053-BB_id.	BG_TR_GR	1_1_2
20060615-1026	SP4	4287726548-DA_4291526585-AB_id.	CH_AT_LI	1_1_2
20060717-1145	IL3	3218236166-DB_3218236166-DB_id.	GB_IE_NI	1_1_1
20060720-1043	IL3	4060529737-AC_4060529737-AC_id.	DE_LU_FR	1_1_1
20060913-0957	IL3	4895829989-BB_4895829989-BB_id.	SK_PL_CZ	1_1_1
20060915-1051	SP4	4834251055-BC_4834251055-BC_id.	SE_FI_NO	2_2_2
20060924-0944	SP4	5154426127-DB_5154426128-BB_id.	CS_RO_HU	1_1_1

(4 SPOT4 and 2 SPOT5) were delivered for both coverages *and* for the very same country (all for either FI or NO). Surprisingly, two of these six duplicates are not identical. The first shows only slight intensity variations but the second shows variations in both intensity and geometry! The large white regions over Northern-Ireland are due to images delivered twice for gb and ni (19 out of 21 ni scenes were also delivered for gb). Rather than compositing, mere selection was applied to (i) the 6 scenes delivered for the same country, (ii) the 19 ni/gb pairs (selecting always the gb image), and (iii) the two cs/me pairs because the current GISCO national territory units version 9 does not distinguish Serbia (cs[3]) from Montenegro (the cs image was arbitrarily selected in both cases). The remaining 130 images that were delivered two or three times were composed using the above described composition procedure. A specific module was required for the 7 duplicates involving Switzerland due to the abnormal cuts in the data ROIs for this country (see Fig. 7). Its description goes beyond the scope of this report.

As a result, 157 unique scenes were generated so as to avoid any duplicates in the European coverage (130 with actual composition and 27 with selection, see above). For all scenes where images were actually composed (no selection), the `countryOrigin`, `parentFile`, and `dataSet` fields of the header file contain a list of 2 (for duplicates) or 3 (for triplicates) values. A graphical representation of the footprints of all composed duplicate scenes is displayed in Fig. 12. The 27 scenes where mere selection occurred are not represented in this figure (i.e., the 19 ni scenes that were also delivered for gb, the 6 duplicates for the same country over the two coverages (4 for fi and 2 for no), and the 2 cs/me duplicates). For the 6 duplicates for the same country over the two coverages, the image from the first coverage has always been selected so that the `dataSet` field is always set to `Image2006Coverage1LAEA` for these images in the reference coverage. It follows that a search for all images with `dataSet` field containing `Image2006Coverage2LAEA` leads to 1,555 images instead of 1,561 as would be expected from table 1.

Figures 11 and 12 also reveal that only a subset of the images overlapping two or more countries were included in all these countries. Consequently, the distinction between the data and reliability domains on the basis of country extent(s) is still kept at the level of the reference coverage. The footprints of all scenes whose CROIs are not equal to their DROIs are displayed in Fig. 13. There are 961 such scenes (333 for SPOT4, 181 for SPOT5, and 447 for IL3).

[3]The ISO code for the republic of Serbia is in fact RS.

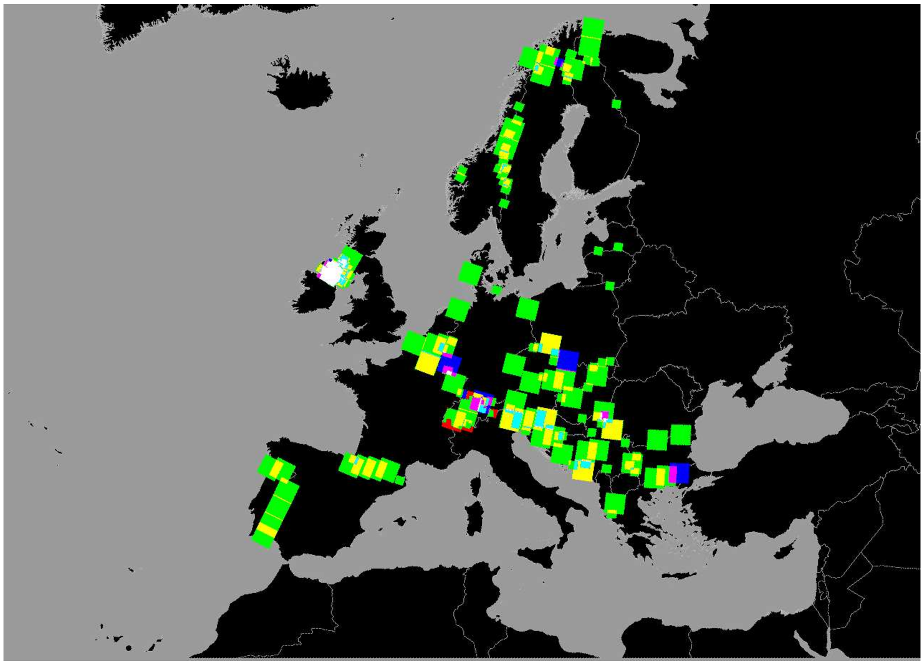


Figure 11: Graphical representation of the footprints of all imagery that was delivered more than once (the colours are used for coding the degree of overlap). See also tables 3 and 4.

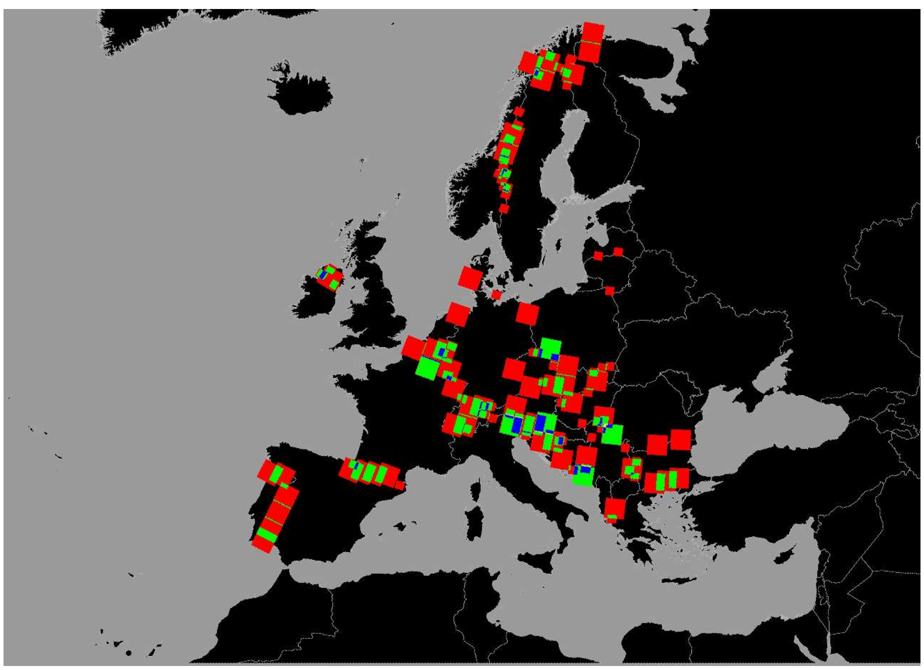


Figure 12: Graphical representation of the footprints of the unique scenes after composition of the duplicate imagery (the 27 scenes where mere selection occurred are not represented). Compare with Fig. 11.

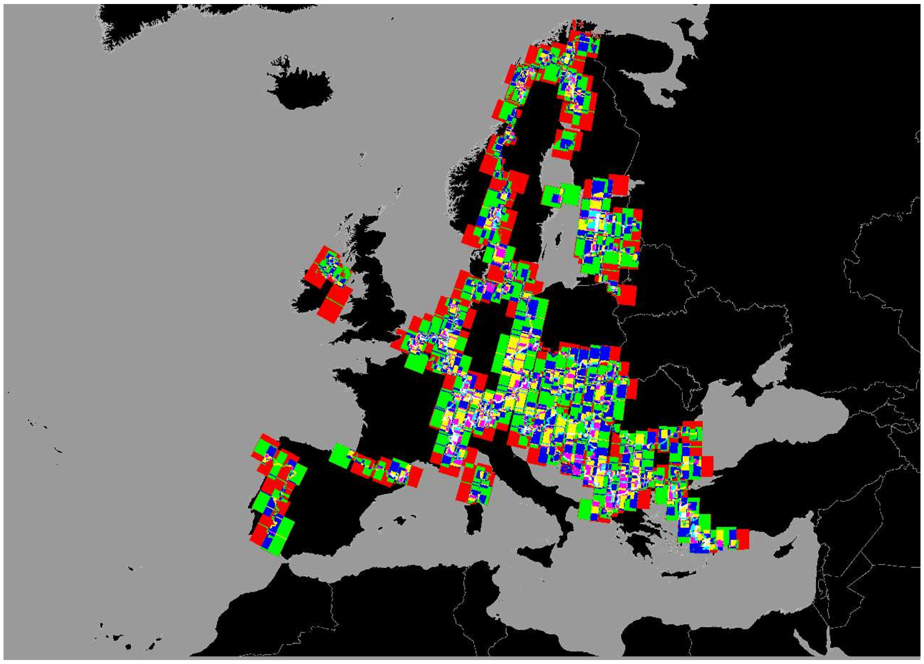


Figure 13: Graphical representation of all 961 CROIs of the reference coverage (total of 3,533 CROIs) that are not equal to their corresponding DROIs.

With the delivery of a truly European coverage, a one-to-one correspondence between raw images and orthorectified images would have been ensured and the reliable area would correspond to the full data footprint of the image (or at least its portion falling within the 38 participating countries).

7 Stored data

All ingested data described in this section is available to the IES SDI and LMNH units at the following location: `netapp2:/geodata/EUIM/`. The data is stored in 3 separate directories (`2006_COV1`, `2006_COV2`, and `2006_REF`), the first 2 for coverages 1 (2080 images) and 2 (1619 images) and the last for the merged coverages with no duplicate imagery (3533 images). Note that `2006_REF` contains symbolic links to the respective files in `2006_COV1` or `2006_COV2` except for the 130 scenes that were composed (see Sec. 6).

7.1 Coverages 1 and 2: `2006_COV1` and `2006_COV2`

The ingested received data for each coverage is stored in the directories named `2006_COV1` and `2006_COV2`. Given that the received coverages can be viewed as a union of non-harmonised country coverages rather than an actual European coverage (see Sec. 1), the identifier string of each generated file name contains the 2 upper case letter country code referring to the lower case country directory the data originated from. For example, `IP-B1ES` translates to band 1 originating from sub-directory named es, i.e., Spain. The full list of country codes used by

DLR is as follows (see caveats in Sec. 1): al at ba be bg ch cs cy cz de dk ee es fi fr gb gr hr hu ie is it li lt lu lv mc me mt ni nl no pl pt ro se si sk tr.

7.2 Combined coverages: 2006_REF

The two input coverages are combined so as to produce a reference data set in the directory named 2006_REF. The identifier string in the reference data set does not contain anymore the 2 letter country code used for coverages 1 and 2. For example, IP-B1ES becomes IP-B1OC (first band, the letter C indicating that cubic convolution was used when interpolating the raw image during orthorectification). Data duplication is avoided for all those 3,376 images that were delivered only once by creating symbolic links to the files already stored in 2006_COV1 or 2006_COV2. Consequently, there are only 130 images physically stored in the directory named 2006_REF since mere selection was considered for 27 (out of a total of 157) duplicate scenes so that links were also produced for them.

References

[1] Adobe Developers Association. TIFF 6.0 specification, 1996. URL http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf.

[2] A. Annoni, editor. *European Reference Grids*, volume EUR 21494 EN. European Commission, DG Joint Research Centre, 2005. URL http://www.ec-gis.org/sdi/publist/pdfs/annoni2005eurgrids.pdf.

[3] A. Annoni, C. Luzet, E. Gubler, and J. Ihde, editors. *Map Projections for Europe*, volume EUR 20120 EN. European Commission, DG Joint Research Centre, 2003. URL http://www.ec-gis.org/sdi/publist/pdfs/annoni-etal2003eur.pdf.

[4] Anonymous. DIMAP Digital Image Map. Technical report, SPOTIMAGE, Toulouse, France, 2008. URL http://www.spotimage.fr/dimap/spec/documentation/refdoc.htm. Last accessed on the 23rd of May 2008.

[5] Anonymous. EPSG Geodetic parameters dataset version.6.15. Technical report, International Association of Oil and Gas Producers, April 2008. URL http://www.ogp.org.uk/.

[6] Anonymous. Computation of at-sensor solar exoatmospheric irradiance and Rayleigh optical thickness for IRS-P6 sensors. Technical report, National Remote Sensing Agency, Bangalore, India, unknown. URL http://www.euromap.de/download/exoatm_1.pdf.

[7] Anonymous. IRS-P6 LGSOWG (Super Structure) Digital Data Products Format. Technical report, Space Applications Centre, ISRO, Ahmedabad, India, May 2003. URL http://www.euromap.de/download/p6super_20050222.pdf.

[8] ISO. ISO 3166 code lists: Country names and codes. International Organization for Standardization, 2008. URL http://www.iso.org/iso/country_codes/iso_3166_code_lists.htm.

[9] A. Meygret. SPOT Absolute calibration: Synthesis. Technical Report S5-NT-0-2880-CN, CNES, Toulouse, France, August 2007.

[10] R. Müller, T. Krauß, M. Lehner, G. Rönnbäck, and Å Karlsson. Image2006 GMES Fast track land service 2006–2008: Orthorectification of SPOT and IRS-P6 products. Technical report, DLR and Metria, May 2008.

[11] M. Pandya, R. Singh, K. Murali, P. Babu, A. Kirankumar, and V. Dadhwal. Bandpass solar exoatmospheric irradiance and Rayleigh optical thickness of sensors on board Indian Remote Sensing Satellites-1B, -1C,-1D, and P4. *IEEE Transactions on Geoscience and Remote Sensing*, 40(3):714–718, March 2002. doi:10.1109/TGRS.2002.1000331.

[12] N. Ritter and M. Ruth. Geotiff format specification revision 1.0, 2000. URL `http://www.remotesensing.org/geotiff/spec/geotiffhome.html`.

[13] J.-F. Rivest, P. Soille, and S. Beucher. Morphological gradients. *Journal of Electronic Imaging*, 2(4):326–336, October 1993. doi:10.1117/12.159642.

[14] P. Soille. IMAGE-2006 mosaic: Composition methodology. Technical report, European Commission, DG Joint Research Centre, 2008. In preparation.

[15] P. Soille. Morphological image compositing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(5):673–683, May 2006. doi:10.1109/TPAMI.2006.99.

[16] P. Soille. IMAGE-2006 mosaic: Relative geometric and radiometric accuracy. Technical report, European Commission, DG Joint Research Centre, 2008. In preparation.

[17] P. Soille and L. Vincent. Determining watersheds in digital pictures via flooding simulations. In M. Kunt, editor, *Visual Communications and Image Processing'90*, volume 1360, pages 240–250, Bellingham, 1990. Society of Photo-Instrumentation Engineers. doi:10.1117/12.24211.

[18] P. Strobl. Organising a shared raster data repository for LMNH and SDI units. Technical report, European Commission, DG Joint Research Centre, 2008.

[19] L. Vincent and P. Soille. Watersheds in digital spaces: an efficient algorithm based on immersion simulations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):583–598, June 1991. doi:10.1109/34.87344.

[20] T. Welch. A technique for high performance data compression. *IEEE Computer Magazine*, 17(6):8–19, June 1984. doi:10.1109/MC.1984.1659158.

European Commission

EUR 23636 EN – Joint Research Centre – Institute for Environment and Sustainability
Title: IMAGE-2006 Mosaic: Data Ingestion and Organisation v1.0
Author: Pierre Soille and Conrad Bielski
Luxembourg: Publications Office of the European Union
2011 – 23 pp. – 21.0 x 29.7 cm
EUR – Scientific and Technical Research series – ISSN 1831-9424 (online), 1018-5593 (print)
ISBN 978-92-79-20958-1
doi:10.2788/50518

Abstract
This report details how the IMAGE-2006 data was ingested and organised in view of creating the IMAGE-2006 mosaic. In particular, it details the method developed for merging the two coverages (delivered on a country basis) into a unique pan-European coverage. The concept of data and country regions of interest is introduced and a method for compositing identical scenes originating from more than one country is detailed. The resulting reference coverage contains 3,533 unique scenes out of a total of 3,699 delivered scenes.

The mission of the JRC is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of EU policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.

JRC
EUROPEAN COMMISSION

LB-ND-23636-EN-N

Publications Office

ISBN 978-92-79-20958-1