GloBAM Data Management Plan

Peter Desmet

2019-04-16

Contents

1 Intro

Welcome to the GloBAM Data Management Plan.

2 European radar data

This chapter describes European radar data (`pvol`) and derived data products (`vp` and `vpts`).

2.1 License agreement

European radar data exchange is coordinated by the *European Operational Program for Exchange of Weather Radar Information (EUMETNET/OPERA)*. GloBAM has access to these data via the *EIG EUMETNET license agreement for Research use of OPERA data*, which grants a specific list of people (involved in ENRAM and GloBAM) access to radar data from selected countries under the following conditions (excerpt):

> The OPERA Members Data will be provided for use within the ENRAM Project for the purpose of extracting animal migration information for scientific research. The Grant of this License does not permit use of the OPERA Members Data licensed to be used for commercial purposes or exploitation for profit.

2.2 Radars

OPERA manages a list of radar metadata (232 radars). This **OPERA database** is available as:

- an interactive map
- a json file: source file for the map. This file is also archived on GitHub every time we update the DMP, so we can track changes.

The license agreement allows access to data from 19 countries:

```
##          country iso_code radars operational
##          Austria       AT      5           5
##          Belgium       BE      3           3
##          Croatia       HR      5           2
##          Czechia       CZ      2           2
##          Denmark       DK      5           5
##          Estonia       EE      2           2
##          Finland       FI     10          10
##           France       FR     31          25
##          Germany       DE     20          20
##      Netherlands       NL      3           2
##           Norway       NO     11          11
##           Poland       PL      8           8
##         Portugal       PT      4           3
##         Slovakia       SK      4           4
##         Slovenia       SI      2           2
##            Spain       ES     15          15
##           Sweden       SE     12          12
##      Switzerland       CH      5           5
##  United Kingdom       GB     16          16
##            Total        -    163         152
```

2.3 PVOL

`pvol` stands for polar volume data. These data are generated by weather radars and are the source for vertical profiles (`vp`) of biological targets.

2.3.1 Format

- Format: `hdf5` in the OPERA ODIM_h5 format.
- Granularity: a `pvol` file is typically generated every 15 minutes and contains scan data at about 20 elevations.
- bioRad function: `read_pvolfile()`.

2.3.2 Source

2.3.2.1 BALTRAD

The best source for consolidated `pvol` data is **BALTRAD**, managed by SMHI (contact person: Günther Haase). `pvol` data are archived there, but access to that server is restricted to SMHI. For data to be available in BALTRAD, several conditions need to be met:

1. A country should send both reflectivity and radial velocity data to the OPERA data centre, which is called ODYSSEY. While many countries are sending reflectivity data, radial velocities are still unavailable for many countries.
2. ODYSSEY should forward these data to the BALTRAD datahub. It is standard policy to do so, but in practice data is not yet forwarded correctly for some radars/countries.
3. Currently BALTRAD and ODYSSEY store data at a 15 minute interval, higher resolution data is not yet available.

2.3.2.2 Research labs

Some research labs (e.g. UvA) also have archived `pvol` data for a subset of radars/years.

2.3.2.3 National weather services

More or more precise `pvol` data can be available at the national weather services as well.

2.3.3 Access

Restricted access: access and use of `pvol` data is subject to the license agreement (2.1). These data cannot be shared further without approval from OPERA. Transferring these data to a (private) cloud platform is also sensitive.

2.3.4 Coverage

Coverage of the `pvol` archive at BALTRAD is currently unknown. Availability and coverage are being gathered in this spreadsheet.

2.3.5 Up-to-dateness

`pvol` date at BALTRAD are normally up to date to within a day.

2.4 PVOL → VP

...

2.5 VP

`vp` stands for vertical profiles of biological targets. These form the main data to be used by GloBAM.

2.5.1 Source

`vp` data generated at BALTRAD are transferred daily to the ENRAM data repository.

2.5.2 Access

Open data: Data in ENRAM data repository are available under a Creative Commons Zero waiver.

2.5.3 Format

- Format: `hdf5` in the ODIM bird profile format specification.
- Granularity: a `vp` file is generated for each source `pvol` file and thus has the same granularity.
- bioRad function: `read_vpfiles()`.

2.5.4 Coverage

The coverage of the ENRAM data repository is recorded daily in `coverage.csv` and summarized here:

```
##  radar 2016 2017 2018 2019 Total
##  bejab   21   NA  194  104   319
##  bewid   19   12  205  105   341
##  bezav   21   NA   NA   NA    21
##  bgvar   21   NA   NA   NA    21
##  ctcdv   21   NA   NA   NA    21
##  ctpda   21   NA   NA   NA    21
##  czbrd   21   NA  286  104   411
##  czska   21   NA  286  105   412
##  deasb   NA   NA  221   93   314
##  deboo   31   17  305   93   446
##  dedrs   31   21  304   93   449
##  deeis   31   21  304   93   449
##  deemd   NA   20   49   NA    69
##  deess   31   21  304   93   449
##  defbg   31   NA   NA   NA    31
##  defld   31   21  303   92   447
##  deflg   NA   21   49   NA    70
##  dehnr   31   21  304   93   449
##  deisn   31   NA   NA   NA    31
##  dememm  31   21  301   93   446
##  deneu   31   21  301   93   446
##  denhb   31   21  300   90   442
##  deoft   31   21  304   93   449
##  depro   31   21  302   93   447
##  deros   25   21  294   93   433
##  desna   NA   21  303   93   417
##  detur   29   21  303   93   446
##  deumd   31   21  304   93   449
##  dkbor   NA   34   85  105   224
##  dkrom   NA   38   85  105   228
##  dksin   NA   38   85  105   228
##  dkste   NA   38   85  105   228
##  dkvir   NA   27   85  105   217
##  eehar   NA   20   35    1    56
##  eesur   NA   20   35   NA    55
##  esalm   NA   38  287  105   430
##  esbad   NA   38  302  105   445
##  esbar   NA   38  303  105   446
##  escor   NA   38  287  105   430
##  eslid   NA   38  303  105   446
##  eslpa   NA   NA   14   NA    14
##  esmad   NA   38  301  105   444
##  esmal   NA   38  303  105   446
##  esmur   NA   38  302  105   445
##  espma   NA   38  303  105   446
##  essan   NA   38  302  105   445
##  essev   NA   36  300  105   441
##  essse   NA   38  303  105   446
##  esval   NA   38  294  105   437
```

```
##  eszar   NA   38  303  102   443
##  fianj   22  154   38   NA   214
##  fiika   22  153   39   NA   214
##  fikes   22  151   38   NA   211
##  fikor   22  146   38   NA   206
##  fikuo   22  156   38   NA   216
##  filuo   22  153   38   NA   213
##  fipet   22   99   NA   NA   121
##  fiuta   22  157   38   NA   217
##  fivan   22  157   39   NA   218
##  fivim   22  132   34   NA   188
##  frabb   21   38  320   95   474
##  frale   21   26  321  105   473
##  frave   21   NA  283  105   409
##  frbla   21   38  321  104   484
##  frbol   21   38  279   99   437
##  frbor   21   36  321  105   483
##  frbou   21   38  173  102   334
##  frcae   21   38  320  105   484
##  frche   21   38  317  105   481
##  frcol   21   38  312  104   475
##  frgre   21   38  320  105   484
##  frlep   21   38  320  100   479
##  frmcl   21   38  304  105   468
##  frmom   21   38  320  105   484
##  frmtc   21   35  302  105   463
##  frnan   21   38  317  105   481
##  frnim   21   38  320  105   484
##  frniz   21   NA  281  105   407
##  fropo   21   38  241   59   359
##  frpla   21   38  315  105   479
##  frtou   21   37  305  105   468
##  frtra   21   38  320  105   484
##  frtre    6   38  316  105   465
##  frtro   21   NA  282  104   407
##  hrbil   NA   38   35   NA    73
##  hrosi   NA   38   35   NA    73
##  nldbl   22    4   NA   NA    26
##  nldhl   21  125  295  105   546
##  nlhrw   NA   21  294  105   420
##  plbrz   21    3  261  105   390
##  plgda   21    3  261  105   390
##  plleg   21    3  259  105   388
##  plpas   NA    3  259  105   367
##  plpoz   21    3  260  105   389
##  plram   21    3  260  105   389
##  plrze   21    3  222  105   351
##  plswi   21    3  260  105   389
##  ptfar  119   NA   NA    4   123
##  ptlis   NA   NA   NA    1     1
##  ptliz  122   NA   NA   NA   122
##  ptprt  113   NA   NA   NA   113
##  seang   21  137  310   94   562
##  searl   21  153  318  103   595
```

```
##  sease   21   16   NA   NA    37
##  sehem   NA   43  314   95   452
##  sehud   21   NA   NA   NA    21
##  sehuv   NA   36  295   97   428
##  sekir   21  144  318  105   588
##  sekkr   21  157  171   NA   349
##  selek   21  118  208   97   444
##  selul   21  148  183   NA   352
##  seoer   NA   36  307   97   440
##  seosd   NA  142  311   95   548
##  seosu   13   NA   NA   NA    13
##  seovi   21   59   NA   NA    80
##  sevar   21   NA   NA   NA    21
##  sevax   NA  133  312   97   542
##  sevil   21  153  318   29   521
##  silis   21   35  304   99   459
##  sipas   21   35  311   69   436
##  skjav   NA   38  301  105   444
##  skkoj   NA   38  301   98   437
##  skkub   NA   NA  280  105   385
##  sklaz   NA   NA  280  105   385
```

2.5.5 Up-to-dateness

The ENRAM data repository is updated daily.

2.6 VPTS

`vpts` stands for time series of vertical profiles. These are a more convenient way of reading `vp` data without any data loss.

2.6.1 Format

Tabular, but to be defined: https://github.com/enram/globam-dmp/issues/16

2.6.2 Source

These files are not yet generated, but will be stored in the ENRAM data repository.

2.6.3 Access

Open access

2.6.4 Coverage

Not yet defined.

2.6.5 Up-to-dateness

Not yet defined.

2.7 Historical archive

GloBAM needs a **2-3 year archive** of European `pvol` and derived `vp` data to tackle its research questions. By starting from a `pvol` archive, the generation of `vp` data can be reproduced if need be, e.g. to make sure adequate vol2bird settings/versions are used.

2.7.1 Source and access

- `pvol`: BALTRAD (2.3.2.1) is likely to be the best source. **Restricted access**
- `vp`: ENRAM data repository (2.5.1). **Open data**
- `vpts`: ENRAM data repository (2.5.1). **Open data**

2.7.2 Coverage

The BALTRAD archive needs to be assessed for coverage and quality before we can define the subset of the archive we want to use. Tasks for this are listed here and include:

1. Get a file listing for `pvol` archive
2. Select subsets: first 2 days of data per radar/month for 2016, 2017, 2018 (72 days): either pvol or pre-merged scans
3. Transfer subsets to accessible FTP server
4. Merge to pvol (test 1)
5. Process with vol2bird to `vp` (test 2)
6. Store output vp files
7. Visual control of `vp` (test 3)