**AMAP biota example: verified test dataset**

**Dataset:**

Time-series of contaminants in tissues and organs of pilot whale (*Globicephala melas*) from the Faroe Islands (Føroyskar hvalvágir sampling location) covering the period 2011-2020.

These data were obtained from the ICES DOME database on 29 August 2023 using the XHAT webservice [https://dome.ices.dk/api/swagger/index.html >>> XHAT >>> XHAT/GetCFDataZIP (and XHAT/GetStationZIP)], applying the criteria country = Denmark; minYear=2011; maxYear=2020; is\_amap\_monitoring=TRUE.

**Dataset citation:**

USFO, 2023. Dataset: Contaminants in pilot whale. Food and Environmental Agency (Heilsufrøðiliga Starvsstovan), Tórshavn, Faroe Islands. Data extracted from: ICES Marine Environmental Database (DOME), Contaminants in Biota. 2023. ICES, Copenhagen, Denmark. <https://dome.ices.dk>. accessed 29 August 2023.

**Data files:**

1. *data/ICES\_DOME\_STATIONS\_20230829.txt* (checksum = XXX)

ICES formatted (tab-delimited .txt) station data file, as delivered by webservice

1. *data/ICES\_FO\_PW\_DATA\_20230829.txt* (checksum = XXX)

ICES formatted (tab-delimited .txt) biota data file; the file delivered by the webservice was filtered to select data for pilot whales from the Føroyskar hvalvágir sampling location in the period 2011-2020. (NB: this was done using software that retained UTF-8 encoding which can be corrupted if e.g. saving tab-delimited files in MS-Excel)

1. *data/EXTERNAL\_AMAP\_STATIONS.csv* (checksum = XXX)

AMAP formatted (comma delimited UTF-8 .csv) station data file

1. *data/EXTERNAL\_FO\_PW\_DATA.csv* (checksum = XXX)

AMAP formatted (comma delimited UTF-8 .csv) biota data file. This file was produced from the ICES formatted biota data file (2) retaining or renaming relevant columns (as below) to comply with the harsat external file format conventions (all other columns were deleted). This approach to preparing an ‘external’ format data file is not necessary, but was applied here to allow direct comparison of harsat results using ICES and external input data.

|  |  |  |
| --- | --- | --- |
| ICES file | External file |  |
| country | country | retained |
|  | station\_code | Inserted (with valid values) |
| statn | station\_name |  |
| myear | year | renamed |
| date | date | renamed |
| latitude | latitude | retained |
| longitude | longitude | retained |
| dephu | dephu | retained (but not required) |
| purpm | purpm | retained (but not required) |
| worms\_accepted\_name | species | renamed |
| sexco | sex | renamed |
| stage | subseries | renamed |
| noinp | n\_individuals | retained (but not required) |
| param | determinand | retained |
| matrx | matrix | retained |
| basis | basis | retained |
| value | value | retained |
| munit | unit | renamed |
| detli | limit\_detection | renamed |
| lmqnt | limit\_quantification | renamed |
| uncrt | uncertainty | renamed |
| metcu | unit\_uncertainty | renamed |
| qflag | censoring | renamed |
| metoa | method\_analysis | retained (but not required) |
| tblsampleid | sample | renamed |

1. *information/determinand.csv* (checksum = XXX)

AMAP ‘determinand’ reference table variant applied in the example runs

1. *information/species.csv* (checksum = XXX)

AMAP ‘species’ reference table variant applied in the example runs

1. *information/threshold.csv* (checksum = XXX)

AMAP ‘threshold’ reference table variant applied in the example runs

**Applied harsat test runs:**

The harsat application was run using the R code below (assume data files are in /data folder and reference tables in /information folder, respectively. Four test cases were run, using ICES and ‘external’ formatted input data, and for ‘subseries’ or no ‘subseries’, respectively.

--------------------------------------------------------------------------------------------

library(harsat)

biota\_data <- read\_data(

compartment = c("biota"),

purpose = c("AMAP"),

contaminants = "ICES\_FO\_PW\_DATA\_20230829.txt",

# NB, replace biota data filename above, as appropriate

stations = "ICES\_DOME\_STATIONS\_20230829.txt",

# NB, replace station data filename above, as appropriate

data\_dir = "data",

data\_format = c("ICES"),

# NB, replace data format option above, as appropriate (external or ICES)

info\_files = list(),

info\_dir = "information",

extraction = NULL,

max\_year = NULL,

oddity\_dir = "oddities",

control = list()

)

# NB, in the above function, replace ‘control = list()’ with ‘control = list(use\_stage = TRUE)’

# to include subseries in ICES format data runs;

# for ‘external’ format runs, if the input data file includes a column named ‘subseries’, the

# data in this column will be used to define subseries, to exclude subseries runs this

# column can be omitted or renamed e.g. to ‘nosubseries’

biota\_data <- tidy\_data(biota\_data)

biota\_timeseries <-create\_timeseries(

biota\_data,

determinands = ctsm\_get\_determinands(biota\_data$info),

determinands.control = NULL,

oddity\_path = "oddities",

return\_early = FALSE,

print\_code\_warnings = FALSE,

get\_basis = get\_basis\_most\_common,

normalise = FALSE,

normalise.control = list()

)

biota\_assessment <- run\_assessment(

biota\_timeseries,

subset = NULL,

AC = NULL,

get\_AC\_fn = NULL,

recent\_trend = 20L,

parallel = FALSE

)

check\_assessment(biota\_assessment, save\_result = FALSE)

write\_summary\_table(

biota\_assessment,

output\_file = "biota-FO-PW-test-output.csv",

# NB, file will be overwritten so change name as appropriate to retain results

output\_dir = "./results",

export = TRUE,

determinandGroups = NULL,

classColour = NULL,

collapse\_AC = NULL

)

# write function writes summary output to /results directory (must exist)

plot\_assessment(

biota\_assessment,

subset = NULL ,

output\_dir = "./results",

file\_type = c("data", "index"),

file\_format = c("png" )

)

# plot function writes output to /results directory (must exist); with these

# settings all plots are output as .png formatted graphics; function can be omitted to

# avoid graphical output; changing “png” to “pdf” in above statement will output graphics

# in PDF (vector) format

--------------------------------------------------------------------------------------------

**Results:**

The above code will write a summary file for the run concerned: Checksums for the four runs described in this example are as follows:

* ICES input data, no subseries = XXXX
* ICES input data run with subseries = XXXX
* external input data, no subseries = XXXX
* external input data run with subseries = XXXX