Empirical Statistical Downscaling

ESD R Package

Rasmus E. Benestad and Abdelkader Mezghani

fou / mk / Tuesday Lunch / Cirrus / 14.01.2014

Main Objectives

- Downscale climate information (variable or parameter) from large (global or regional) to local scales (station),
- Empirical-statistical relationships between a set of predictands and predictors,
- Free package based on R programming language,
- Quick statistical analysis
- Tailored package for different users,
- Flexible
- Traceability of the results and methods used
- ...

How “esd” is relevant for the RCM community ?

Can make use of esd for the RCM community :

- Apply “esd” to RCM results,
- quick look at the RCM results (mapping, plotting,...)
- quick statistical analysis, e.g.
  - Canonical correlation analysis,
  - Trend analysis,
- Compare ESD to RCM results,
- Other diagnostics and validations

Facilitating comparisons

Quality of the data.

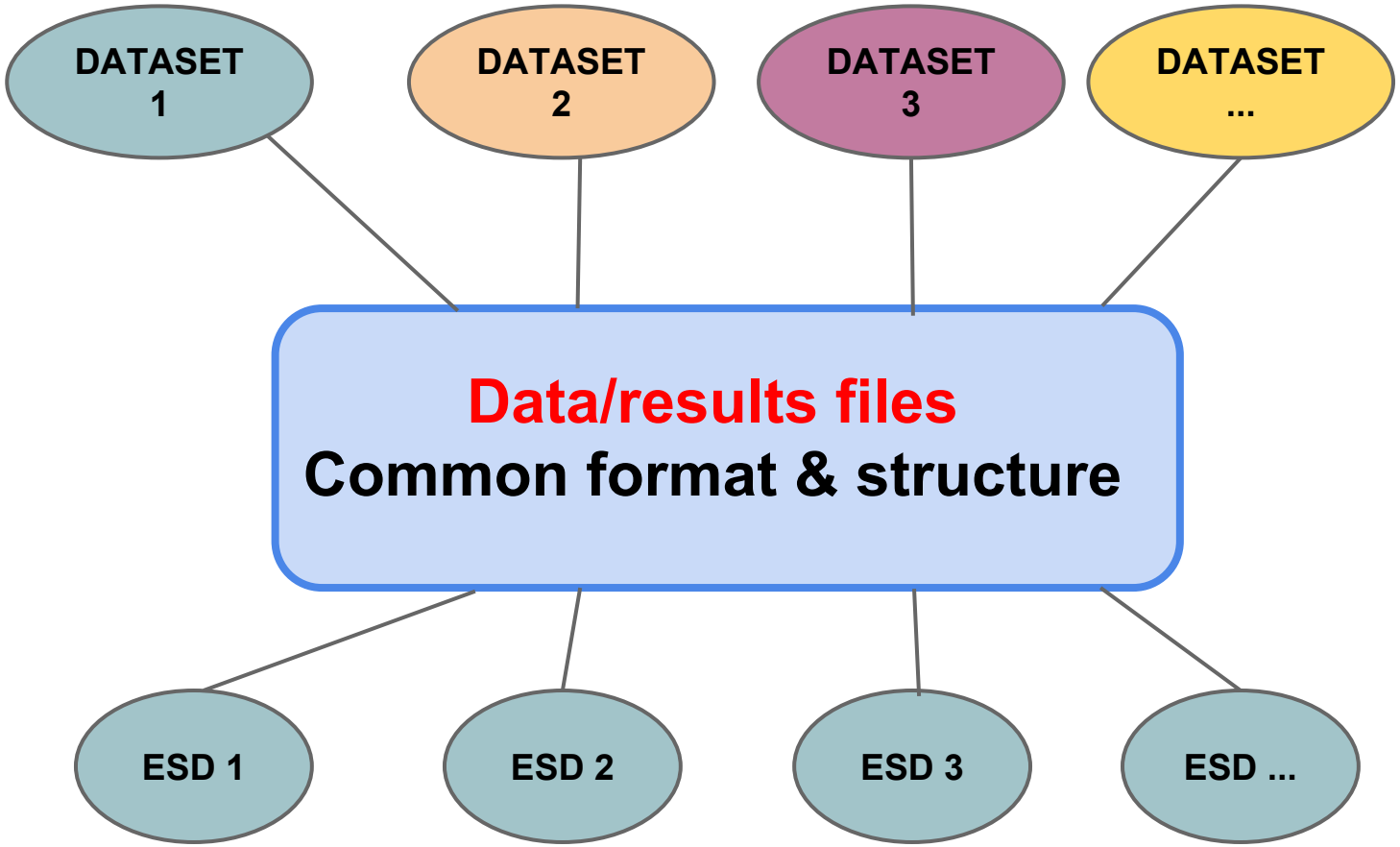
Attributes enhancing traceability

Conventions, standards, & attributes

(CF, netCDF, CMIP)

e.g. make users expect common ‘history’, ‘quality’, ‘method-test’ attributes, describing e.g. efforts in method evaluation (cross-validation, out-of-sample test, blind testing).

Example



Evaluate and Compare !

“esd” R package

- S3 Classes and Methods

  i.e. plot(x) and plot.station(x) are equivalent if x is an object of type “station”

- based on "zoo" class

  S3 Class and methods for indexed totally Ordered observations (ordered values for regular (ts) and irregular time series (zoo))

- installation

  from ftp://ftp.met.no/users/rasmusb/ (# not yet available on CRAN)

  download the latest version esd_0.5-5.tar.gz

  $ R CMD INSTALL esd_0.5-5.tar.gz”

  > install.packages(“esd_0.5-5.tar.gz”)

- Load the library

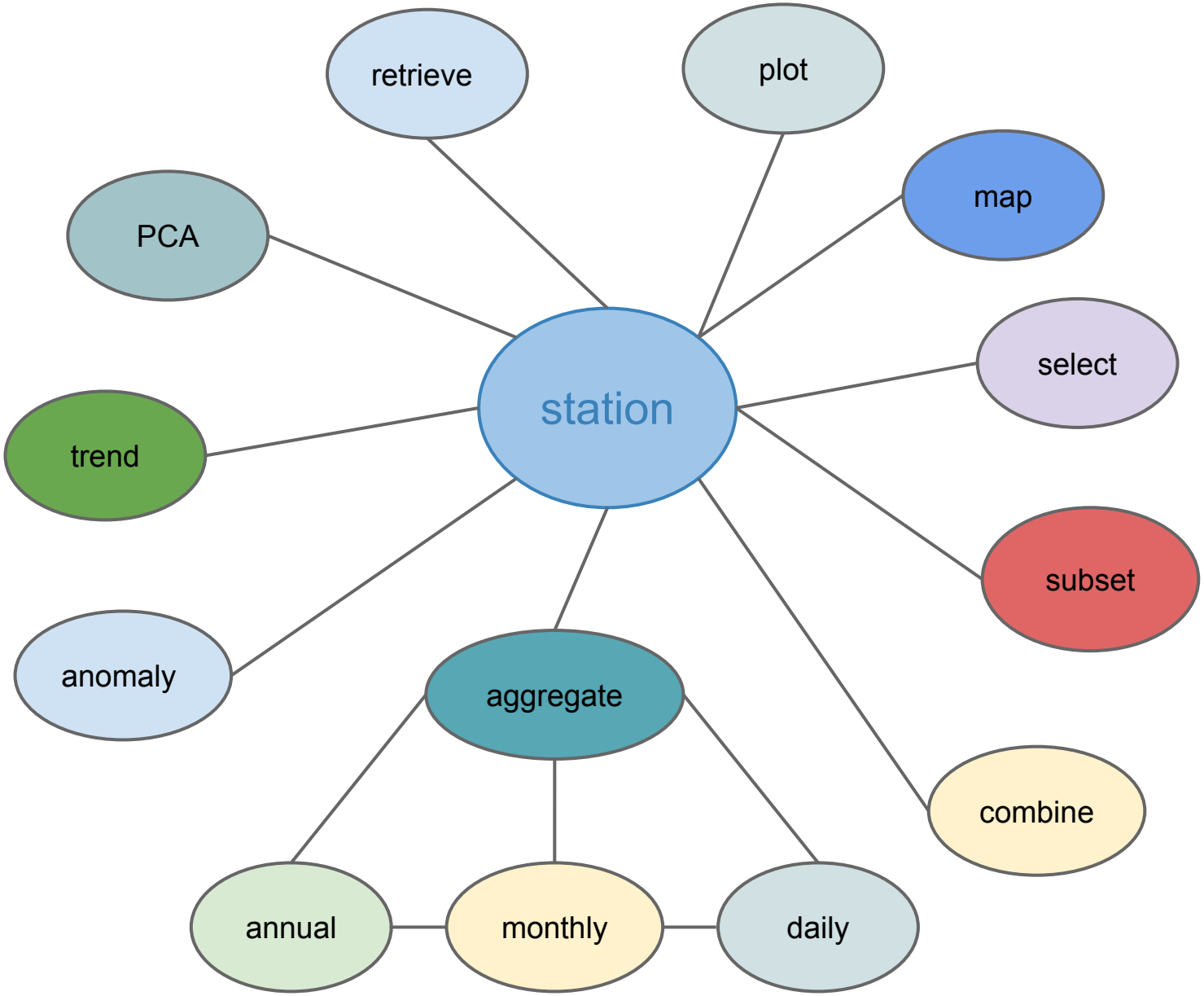
  > library(esd)

- User guide “esd.pdf”

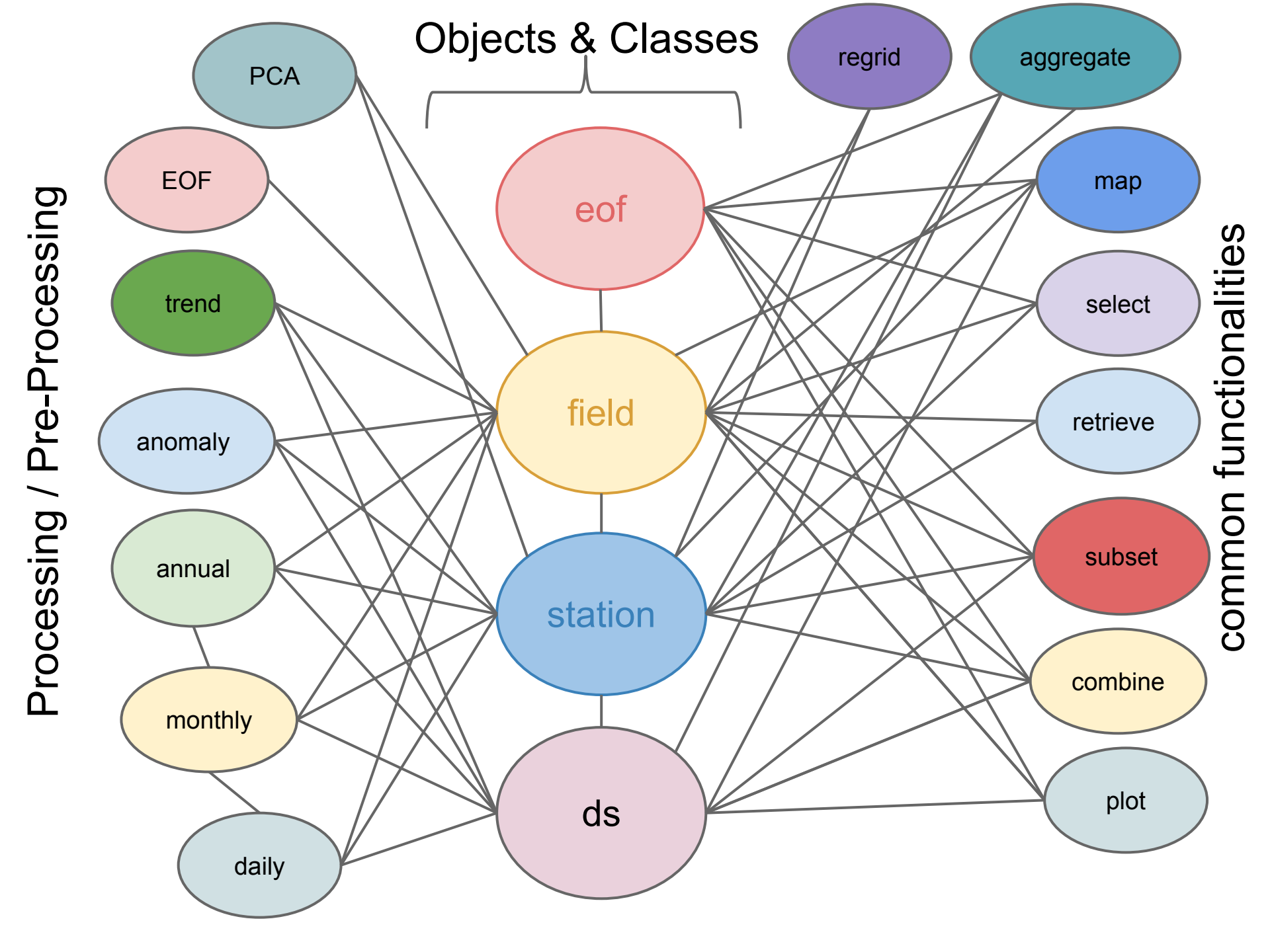
  ftp://ftp.met.no/users/rasmusb/esd.pdf

Objects & Classes

Processing / Pre-Processing



common functionalities


Objects & Classes
PCA
EOF
trend
anomaly
annual
monthly
daily
eof
field
station
ds
regrid
aggregate
map
select
retrieve
subset
combine
plot
Processing / Pre-Processing
common functionalities

Attributes

Common attributes:

‘station_id’, ‘country’, ‘longitude’, ‘latitude’, ‘altitude’, ‘URL’, ... ‘parameter’,
‘longname’, ‘unit’, …
‘timeunit’, ‘calendar’,‘frequency’, ...
‘source’, ‘aspect’, ‘quality’,‘reference’, ‘info’, ...
‘experiment’, ‘model’, and ‘realization’, …
‘method’ , ...
‘history’, ‘filename’, ...

More specific attributes

‘dimensions’ , ‘pattern’ , ‘fitted_values’, …

Additional attributes defined by user needs

Main functionalities / DATA handling

- data(package=”esd”)
- station() → for weather stations
- retrieve() → for any field e.g. Reanalysis, GCMs, ...

output : a “zoo” “station” or “field” object with attributes

Data sets in package ‘esd’:

NACD	Sample data.	eof.precip.ERAINT	Sample data.	global.t2m.cmip3	
NAOI	Sample data.	eof.slp.DNMI	Sample data.	global.t2m.cmip5	
NARP	Sample data.	eof.slp.ERAINT	Sample data.	met.no.meta	
NINO3.4	Sample data.	eof.slp.MERRA	Sample data.	nordklim.data	
Oslo	Sample data.	eof.slp.NCEP	Sample data.	precip.NORDKLIM	Sample data.
Svalbard	Sample data.	eof.sst.DNMI	Sample data.	scandinavia.t2m.cmip3	
arctic.t2m.cmip3		eof.sst.NCEP	Sample data.	scandinavia.t2m.cmip5	
arctic.t2m.cmip5		eof.t2m.DNMI	Sample data.	station.meta	Sample data.
bjornholt	Sample data.	eof.t2m.ERA40	Sample data.	sunspots	Sample data.
etopo5		eof.t2m.ERAINT	Sample data.	t2m.NORDKLIM	Sample data.
ferder	Sample data.	eof.t2m.MERRA	Sample data.	vardo	Sample data.
geoborders	Sample data.	eof.t2m.NCEP	Sample data.		
		eof.t2m.NorESM.M	Sample data.		

Main functionalities / Processing

- select.station() , subset() , aggregate() , ...
- EOF()
- regrid()

  regrid.eof() , regrid.field() , ...
- combine()

  combine.eof() , combine.field() , combine.station() , ...
- DS()
- result to a new object belonging to appropriate classes with updated + new attributes

ASPECT ‘anomaly’, ‘climatology’, ‘pattern’, ‘group-of-stations’ , ...

- plot() , map()

  plot.station() , plot.field() , plot.eof() , plot.ds() , ... , map.eof() , map.field() , ...

Benefits of common standards & structures

- Facilitate implementations
- Ease intercomparisons
- Sharing of generic methods
- Traceability and replicability
- Promotes community building
- Promotes discussions

ESD framework in 3 steps !

1. Select and process
   - Station(s) : e.g Oslo, Norway, Scandinavia, Europe, ...
   - Parameter(s) : t2m, precip, …
   - Predictor : global air temperature, Sea level pressure, ...
   - Reanalysis : t2m.ERAINT, t2m.MERRA, …

2. DS strategy and methods
   - Method : lm, glm, …
   - Strategy : e.g. DS of EOF of station then reconstruct

3. Plot results and diagnostics

Data objects

Primary

‘station’ - Single or Multivariate time series of observations (stations)

‘field’ - Time series of gridded values (model results, analyses)

‘ds’ - Time series of downscaled values (ds results, analyses)

Secondary

EOFs

PCAs

CCAs

Diagnostics

Structures

- S3 methods and “zoo” class objects
- Most sensible ways of representing sets of stations, fields, EOFs,etc.
- Distinguish observations from predictions.
- Based on common R methods: plot(), map(), aggregate(), print(), predict(), ...

Object “station”

Retrieve the data for “Oslo-Blindern” (stid=”193”) from the ECA&D dataset

> obs <- station(loc="oslo blindern", stid="193",src="ecad") or

> obs <- station.ecad(loc="oslo blindern", stid="193")

[1] "Retrieving data ..."

[1] "1 T2M 193 OSLO BLINDERN NORWAY ECAD"

> str(obs)

‘zoo’ series from 1937-03-01 to 2013-08-31

Data: atomic [1:27943] 1.5 1.5 0 -1.6 -4.6 -4.9 -8.9 -9.2 -9.8 -8.9 ...

- attr(*, "location")= chr "OSLO BLINDERN"
- attr(*, "variable")= chr "t2m"
- attr(*, "unit")= chr "degree Celsius"
- attr(*, "longitude")= num 10.7
- attr(*, "latitude")= num 59.9
- attr(*, "altitude")= num 94
- attr(*, "country")= chr "NORWAY"
- attr(*, "longname")= chr "Mean temperature"
- attr(*, "station_id")= chr "000193"
- attr(*, "quality")= int NA
- attr(*, "calendar")= chr "gregorian"

….

- attr(*, "source")= chr "ECAD"
- attr(*, "URL")= chr "http://eca.knmi.nl/utils/downloadfile.php?file=download/ECA_blend_all.zip"
- attr(*, "type")= logi NA
- attr(*, "aspect")= chr "original"
- attr(*, "reference")= chr "Klein Tank, A.M.G. and Coauthors, 2002. Daily dataset of 20th-century surface air temperature and precipitation series for the "| __truncated__
- attr(*, "info")= chr "Data and metadata available at http://eca.knmi.nl"
- attr(*, "method")= logi NA

...

Index: Date[1:27943], format: "1937-03-01" "1937-03-02" "1937-03-03" "1937-03-04" ...

> class(obs)

[1] "station" "month" "zoo"

- attr(*, "history")=List of 3
  ..$ call :List of 1
  .. ..$: language ecad.station(stid = stid[i], lon = lon[i], lat = lat[i], alt = alt[i], loc = loc[i], cntr = cntr[i], qual = qual[i], param = param[i], verbose = verbose, ...
  ..$ timestamp: chr "Tue Sep 9 15:58:44 2014"
  ..$ session :List of 3
  .. ..$ R.version : chr "R version 3.0.3 (2014-03-06)"
  .. ..$ esd.version: chr "esd_0.5-4"
  .. ..$ platform : chr "x86_64-pc-linux-gnu (64-bit)"

Object “field”

Retrieve the 2-m temperature time series from ERA40 datasets

> era40 <- retrieve(ncfile=“data/t2m.era40.mon.nc”,lon=NULL , lat=NULL)

> class(era40)

[1] "field" "month" "zoo"

> str(era40)

time

‘zoo’ series from 1957-09-01 to 2002-08-01

Data: num [1:540, 1:16380] -62.9 -51.4 -34.2 -20.3 -21.9 ...

- attr(*, "variable")= chr "t2m"
- attr(*, "longname")= chr "temperature at 2m"
- attr(*, "unit")= chr "deg C"
- attr(*, "source")= chr "ERA40"

longitude latitude time

- attr(*, "dimensions")= int [1:3] 180 91 540
- attr(*, "longitude")= num [1:180(1d)] 0 2 4 6 8 10 12 14 16 18 ...
- attr(*, "latitude")= num [1:91(1d)] -90 -88 -86 -84 -82 -80 -78 -76 -74 -72 ...
- attr(*, "greenwich")= logi TRUE
- attr(*, "calendar")= chr "gregorian"
- attr(*, "type")= logi NA
- attr(*, "aspect")= chr "original"

….

….

….

….

Index: Date[1:540], format: "1957-09-01" "1957-10-01" "1957-11-01" "1957-12-01" ...

- attr(*, "history")=List of 3

..$ call :List of 3

.. ..$: language eof2field(eof.t2m.ERA40, lon = lon, lat = lat, anomaly = anomaly)

.. ..$:length 19 as.field(t(t2m.in), index = as.Date (tim), lon = lon, lat = lat, param = "t2m", unit = "deg C", alt = NA, loc = NA, cntr = NA, longname = "temperature at 2m", ...

.. .. ..- attr(*, "srcref")=Class 'srcref' atomic [1:8] 353 1 356 81 1 81 353 356

.. ..- attr(*, "srcfile")=Classes 'srcfilecopy', 'srcfile' <environment: 0x422cda8>

.. ..$: chr "unknown past"

..$ timestamp: chr [1:3] "Wed Sep 10 09:23:16 2014" "Thu Dec 5 09:02:12 2013" "unknown past"

..$ session :List of 3

.. ..$ R.version : chr "R version 3.0.3 (2014-03-06)"

.. ..$ esd.version: chr "esd_0.5-4"

.. ..$ platform : chr "x86_64-pc-linux-gnu (64-

Object “eof”

Compute the EOFs for january (it=1) from the ERA40 datasets

> eof <- EOF(era40,it=1)

> class(eof)

[1] "eof" "field" "month" "zoo"

> str(eof)

time

‘zoo’ series from 1958-01-01 to 2002-01-01

Data: num [1:45, 1:20] -0.0707 0.0951 -0.1054 0.0313 0.1447 ...

- attr(*, "variable")= chr "t2m"
- attr(*, "longname")= chr "temperature at 2m"
- attr(*, "unit")= chr "deg C"
- attr(*, "source")= chr "ERA40"
- attr(*, "longitude")= num [1:180(1d)] 0 2 4 6 8 10 12 14 16 18 ...
- attr(*, "latitude")= num [1:91(1d)] -90 -88 -86 -84 -82 -80 -78 -76 -74 -72 ...
- attr(*, "greenwich")= logi TRUE
- attr(*, "calendar")= chr "gregorian"
- attr(*, "type")= logi NA
- attr(*, "aspect")= chr "anomaly"
- attr(*, "dimnames")=List of 2
  ..$: NULL
  ..$: chr [1:20] "X.1" "X.2" "X.3" "X.4" ...

longitude latitude Number of eofs

- attr(*, "pattern")= num [1:180, 1:91, 1:20] -0.00117 -0.00117 -0.00117
- attr(*, "dimensions")= int [1:3] 180 91 45
- attr(*, "mean")= num [1:180, 1:91] -22.7 -22.7 -22.7 -22.7 -22.7 ...
- attr(*, "max.autocor")= num 0.997
- attr(*, "eigenvalues")= num [1:20] 789 652 498 419 352 ...
- attr(*, "sum.eigenv")= num 4934
- attr(*, "tot.var")= num 2051000
- attr(*, "area.mean.expl")= logi FALSE
  Index: Date[1:45], format: "1958-01-01" "1959-01-01" "1960-01-01" "1961-01-01" ...

- attr(*, "history")=List of 3
  ..$ call :List of 4
  .. ..$: language EOF.field(era40, it = 1)
  .. ..$: language eof2field(eof.t2m.ERA40, lon = lon, lat = lat, anomaly = anomaly)
  .. ..$:length 19 as.field(t(t2m.in), index = as.Date(tim), lon = lon, lat = lat, param = "t2m", unit = "deg C", alt = NA, loc = NA, cntr = NA, longname = "temperature at 2m", ...
  .. .. ..- attr(*, "srcref")=Class 'srcref' atomic [1:8] 353 1 356 81 1 81 353 356
  .. ..- attr(*, "srcfile")=Classes 'srcfilecopy', 'srcfile' <environment: 0x422cda8>
  .. ..$: chr "unknown past"
  ..$ timestamp: chr [1:4] "Wed Sep 10 10:38:38 2014" "Wed Sep 10 09:23:16 2014" "Thu Dec 5 09:02:12 2013" "unknown past"
  ..$ session :List of 3
  .. ..$ R.version : chr "R version 3.0.3 (2014-03-06)"
  .. ..$ esd.version: chr "esd_0.5-4"
  .. ..$ platform   : chr "x86_64-pc-linux-gnu (64-bit)"

Object “ds”

```
# Downscale for january (it=1) from the ERA40 datasets
ds <- DS(y1,eof)
> class(ds)
[1] "ds"       "eof"   "field" "month" "zoo"
> eof <- EOF(era40,it=1)
> class(eof)
[1] "eof"   "field" "month" "zoo"
> names(attributes(ds))
 [1] "index"              "class"              "names"
 [4] "location"           "country"            "station_id"
 [7] "longitude"          "latitude"           "altitude"
[10] "variable"           "longname"           "unit"
[13] "aspect"             "source"             "quality"
[16] "URL"                "history"            "reference"
[19] "info"               "calibration_data"   "fitted_values"
[22] "original_data"      "model"              "mean"
[25] "method"             "eof"                "pattern"
[28] "dimensions"         "type"               "history.predictand"
[31] "evaluation"
```

```
> attr(ds,"method")
[1] "lm"
> attr(ds,"model")

Call:
lm(formula = y ~ X.2 + X.3 + X.4 + X.5 + X.7, data = caldat)

Coefficients:
(Intercept)         X.2         X.3         X.4         X.5         X.7
  1.645e-17  -1.161e+01           4.252e+00           4.502e+00  -2.724e+00 4.643e+00
```

```
> attr(ds,"fitted_values")
 1958-01-01  1959-01-01  1960-01-01  1961-01-01  1962-01-01  1963-01-01
 1.72129518  0.80596745  1.09172172  1.00555065  3.45002628 -2.17537235
 1964-01-01  1965-01-01  1966-01-01  1967-01-01  1968-01-01  1969-01-01
 1.85486775  2.36552724 -0.15814646 -1.25234297  1.69375138 -2.03425612
 1970-01-01  1971-01-01  1972-01-01  1973-01-01  1974-01-01  1975-01-01
 1.06200590  0.94740308  1.53442605  2.73348021  2.78645794  4.39370857
 1976-01-01  1977-01-01  1978-01-01  1979-01-01  1980-01-01  1981-01-01
 1.60426977 -1.20515839 -0.51313709 -4.81018869 -0.89916576  0.23723186
 1982-01-01  1983-01-01  1984-01-01  1985-01-01  1986-01-01  1987-01-01
-2.20610137  3.04856225  2.89733074 -4.14695740 -1.07088809 -2.97229144
 1988-01-01  1989-01-01  1990-01-01  1991-01-01  1992-01-01  1993-01-01
 0.26779205  3.88018815  2.82753537  2.22103851  0.99054260  2.32996270
 1994-01-01  1995-01-01  1996-01-01  1997-01-01  1998-01-01  1999-01-01
-0.73704702  3.29560761 -0.16275666 -1.62488122  0.03509849  0.84182816
 2000-01-01  2001-01-01  2002-01-01
 3.28358115  0.81350331  1.94842889
```

ESD tools

esd: man-pages, R-scripts, data, examples

esd - user's guide

- esd.pdf (documentation)

esd - main methods

- station
- EOF, PCA
- DS

esd - data

- Observational networks → ECAD, GHCN, METNO, …
- Reanalysis datasets → NCEP, ERA (40, INTERIM), MERRA, JRA55, …
  GCMs → CMIP3, CMIP5 experiments
  Predefined datasets (temperature, precipitation, global mean temperature,...)

esd - examples DEMO

Summary

- R ESD open source package
- Predefined sets of data and downscaling methods and strategies
- Results traceability
- Feedbacks and updates on a Facebook page (https://www.facebook.com/Rclimateanalysis)
- Flexible tool tailored for several user's need : research, academic, ...