



Seventh Framework Programme
Theme 6 [SPACE]



Project: 607193 UERRA

Full project title:
Uncertainties in Ensembles of Regional Re-Analyses

Deliverable D3.3:
**A portable starting set of algorithms to
support the common evaluation procedure**

WP no:	3
WP leader:	DWD
Lead beneficiary for deliverable :	DWD
Name of <u>author</u> /contributors:	<u>Michael Borsche, Cristian Lussana, Andrea Kaiser-Weiss</u>
Nature:	Other
Dissemination level:	PU
Deliverable month:	15
Submission date: June 19, 2015	1 st Version shared via email 19 June 2015



Report for Deliverable 3.3 (D3.3): A portable starting set of algorithms to support the common evaluation procedure, from UERRA WP3

By Michael Borsche¹, Cristian Lussana², and Andrea Kaiser-Weiss¹

¹ Deutscher Wetterdienst (DWD), Offenbach, Germany

² Norwegian Meteorological Institute (Metno), Oslo, Norway

1. Programme package shared

As a result of WP3 work (Metno and DWD), a program package was shared with all UERRA partners with an email on 19th June 2015 11:44:

Dear UERRA partners,

I would like to announce the starting set of reanalysis evaluation routines (=WP3 deliverable D3.3) which was developed by Cristian Lussana and Michael Borsche. It can be accessed via <https://github.com/UERRA-EVA>

Here, R-routines are provided which compare reanalysis against gridded data: https://github.com/UERRA-EVA/EVA_gridobs

and an R-package comparing reanalysis against station data: https://github.com/UERRA-EVA/EVA_stationobs

together with the test data, or access to test data, respectively. More documentation on the scope of the R-routines, and on how to install and use, can be found directly there.

Current focus is on wind speed (from station data of DWD) and on precipitation (gridded data of Met.no). Time series, anomalies and correlations are compared and pdf-related skill scores implemented. The tests now use ERA-20C, ERA-Interim, ERA-Interim downscaling (NORA10) and COSMO-HErZ, both of the latter because they are of easy access and natural interest for DWD and Met.no.

Of course, in future, the R-routines will draw data from the UERRA archive, and will be applicable to all regional analyses produced in UERRA. Such input routines are in work, but will be more useful later, when the UERRA archive is filled up. Likewise, the independent data sets produced in UERRA can be included when ready.

<https://github.com/UERRA-EVA> will hold the working repositories for WP3, i.e., code actively developed further and more skill scores will be added during the UERRA project. If you have R-packages to add - this will apply mainly to the WP3 partners, but everybody is welcome - , please contact Michael or Cristian.



2. R-package and R-Routines

This starting set of algorithms has been programmed in R and is explained in more detail below. It contains installation instructions and sample data.

The package **EVA_stationobs** is the UERRA common evaluation procedure for assessing uncertainties in reanalysis by evaluation against station observations.

The collection of R-scripts **EVA_gridobs** is the UERRA common evaluation procedure: assessing uncertainties in reanalysis by evaluation against gridded observational datasets.

EVA_stationobs

This package uses German station observations to compare against global (ERA20C and ERA-Interim) and regional (COSMO HERZ) reanalyses by plotting and analysing station and co-located reanalysis time series. The verification so far is performed on monthly means.

Station observations are downloaded by the package during run-time from the CDC-ftp server of DWD. The reanalysis data had been downloaded from ECMWF (global reanalyses) and directly from Uni Bonn (regional reanalysis) and post processed to a time series of monthly means. These data files are necessary for the package to run properly. The post-processed files can be accessed through ECMWF's server ecgate at /scratch/ms/de/ded3/data/EVA_stationobs/.

Further data sources may be added later, i.e., other regional reanalyses or station observations when or if available.

So far, this package (EVAstatobsR) is distributed as a source package, i.e., the pure source code is provided. With the package "devtools" it can be installed; RStudio provides easy access to build functionality. The folder EVA_statobs provides the environment under which the package is useable. It includes:

- the run script AnalyseWindspeed.R and its configuration file Settings.R
- folders from which data is read and into which data and results will be written during run-time
 - ./data/, ./DWD_station_data/, ./output/

The verification can be run on four different pre-defined time periods. For each time period, four different analyses can be performed:

- plot monthly mean time series of station data and reanalyses, their anomalies, and correlations of 10m wind speed
- the same as above for 100m wind speed between ERA20C and COSMO HERZ reanalyses
- the same as the first for specific monthss only
- calculate an S_score between two PDFs and determine their overlapping following Mayer et al., 2015 (determine whether the PDFs belong to the same distribution with the Smirnov test will be added soon)



EVA_gridobs

Reanalysis and gridded observations are assumed to be: on the same coordinate reference system and grid; same temporal aggregation.

Tip: use fimex for regridding (<https://wiki.met.no/fimex/start>)

List of Skill-Scores/Tests:

- Developed and tested for daily precipitation
- Probability Density Functions (PDFs) related skill-scores (PDFs are approximated by comparing discrete histograms):
 - difference between modes: $\text{mode}(\text{reanalysis}) - \text{mode}(\text{observation})$
 - relative precipitation biases = $(\text{mode}(\text{reanalysis}) / \text{mode}(\text{observation}) - 1) * 100\%$
 - overlapping skill-score (see [1], Eq.(1))
- two-sample Kolmogorov-Smirnov (K-S) test, or Smirnov test (see [2], Eqs (5.17-18))
- Fractional skill-score
- Optical-flow

[1] Mayer, S. et al. (2015). Identifying added value in high-resolution climate simulations over Scandinavia. Tellus A

[2] Wilks, D. S. (2011). Statistical methods in the atmospheric sciences (Vol. 100). Academic press.

Running the programs (examples):

1. Edit configuration file according to the instructions reported in `.../etc/config_list.r`
2. Create histograms approximating the frequency distribution of precipitation values (note: see in the directory "data" for test data with frequency distributions)

for the reanalysis

```
R --vanilla your_configuration_file < prec_freqhist_reanalysis.R
```

for the gridded observational dataset (is it possible to aggregate the data to a coarser resolution)

```
R --vanilla your_configuration_file < prec_freqhist_gridobs.R
```

3. Get the mode of reanalysis and gridded observation PDF:

```
R --vanilla your_configuration_file < mode.R
```

4. Compute the difference between modes and relative precipitation biases:

```
R --vanilla your_configuration_file < modediff.R
```



```
R --vanilla your_configuration_file < modeRelErrscores.R
```

5. Compute the PDF overlapping skill-score:

```
R --vanilla your_configuration_file < pdfscores.R
```

6.K-S test:

```
R --vanilla your_configuration_file < KScores.R
```

3. Summary and Outlook

This is a description of ongoing code development. More scores will be added during the run of WP3. Adapters to the UERRA Archive will be in place as soon as the archive is populated.

Please refer to the on-line resource <https://github.com/UERRA-EVA> for the latest results.