Seventh Framework Programme
Theme 6 [SPACE]

Project: 607193 UERRA

Full project title:
Uncertainties in **E**nsembles of **R**egional **R**e-**A**nalyses

Deliverable D8.4
User Guidance

WP no.:	7
WP leader:	KNMI
Lead beneficiary for deliverable:	KNMI
Name of author/contributors/review:	Gé Verver
Nature:	Other
Dissemination level:	PU
Deliverable month:	48
Submission date:	18 December 2017
Version nr:	1

Table of Contents

1. Introduction

Reanalysis integrates and synthesizes observations from satellites, ground-based stations and results from comprehensive models into a consistent dataset describing the state and historical evolution of the climate. The UERRA-project (UERRA: Uncertainties in Ensembles of Regional Reanalyses; www.uerra.eu) produces long-term reanalyses datasets of Essential Climate Variables (ECVs) at the European regional scale in support of adaptation action and policy development and contribute to climate services, climate monitoring and research.

One of the key objectives of the project is to provide an ensemble of reanalyses results to allow for an assessment of the uncertainty. Ensembles are obtained by generating multiple results from a single reanalyses system, as well as by using a number of different reanalyses systems. Improved gridded datasets, based on in-situ observations only, were developed for validation.

The assessment and quantification of uncertainties is crucial for the interpretation of the reanalysis products and their proper use in applications and downstream services. In UERRA, the information content of the regional reanalyses and their uncertainties are statistically assessed by comparison against independent or different ECV datasets at the user relevant scales. The reference datasets include satellite-derived data for climate monitoring, and gridded datasets based on high-density station series together with their uncertainty estimates. A report describing the detailed assessment is available on the UERRA website (www.uerra.eu). The website also provides a user-friendly synthesis report on evaluation and uncertainty of regional reanalyses. In this guide, therefore only a brief summary of these will be provided.

Input for this short User Guide was provided by two user workshops, one in Toulouse, France, on 3-4 February 2016, focussing on user requirements, and one in Tarragona, Spain, on the 30th of November 2017, focussing on the UERRA datasets. More information on these workshops, including the presentations can be found on the project website.

It is clear that this guide cannot cover the use of reanalyses data for all possible user applications in much detail. Although UERRA provides consistent datasets describing the state and historical evolution of the European climate, there are strengths and weaknesses inherent to reanalyses in general, and sometimes specific to a certain reanalyses system. The assessment of whether the UERRA dataset is ‘fit-for-purpose’ has to be made by the user, who has the knowledge of the requirements and sensitivity of the downstream application for the (quality of the) climate data that are available.

This guide also provides a short outline on how to access and use reanalysis data in climate services. More information on this topic can be found under “Training Material” on the UERRA website.

2. Principles of reanalysis systems

Atmospheric reanalysis is a method to reconstruct the past weather by combining historical observations (in situ, surface and satellite remote sensing), with a dynamical model. It provides a physically and dynamically coherent description of the state of the atmosphere. Typically, this is done in a stepwise fashion, where a model provides a model forecast, which is then corrected on the basis of the observations (see *figure 1*). This corrective step, referred to as 'data assimilation', requires statistical knowledge of the forecast error and the observation error. Reanalysis systems differ in the set of observations that is assimilated, the model that is used, and the way the error statistics are estimated and corrections are applied.

A large variety of reanalysis methods exist, and only some of them are used in the UERRA project. Below we provide an explanation of some of the main concepts required for understanding reanalysis data.

3D-VAR, *4D-VAR*, *and nudging*
A reanalysis system as depicted in *Figure 1* is called '3DVar': at fixed points in time the model state is adjusted based on the observed state, taking into account the statistics of model and observation errors. A reanalysis system that considers observations distributed within a time interval of, for instance, 12 hours to take into account the flow-dependent errors, is labeled '4DVar'.
'Nudging' is a technique to distribute the adjustments over all time steps, thus keeping the model state closer to the observations at all times.

Ensemble vs Deterministic
Some reanalysis systems used in UERRA generate an ensemble of results. The different ensemble members are generated by adding a random perturbation in the model and/or in the observations. From the ensemble results, the ensemble mean is often used as the best estimate of the 'truth', and the ensemble spread is used as a measure of the uncertainty of this estimate. Other systems, referred to as 'deterministic', produce one single optimal estimate of the state of the atmosphere.
A hybrid approach is also used, where an ensemble model forecast is used only to estimate the forecast error (the blue dots in *figure 1*), which is subsequently used to generate a single optimal solution.

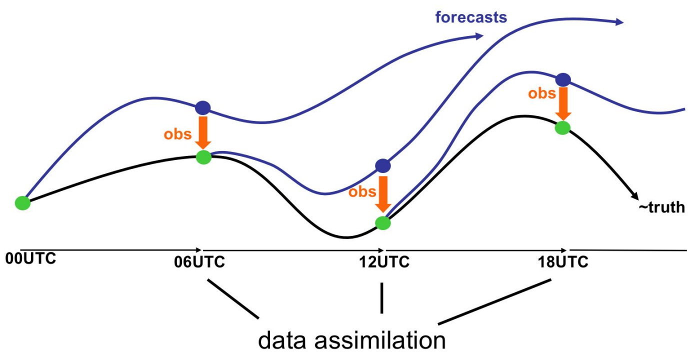

Figure 1: Model forecasts (blue lines) and a stepwise assimilation of observations to obtain an optimal estimate (black line) of the state of the atmosphere. (Figure by Richard Renshaw, UK MetOffice)

The main advantages of reanalyses are:

- They provide regularly gridded data, even in places where there are no or few observations;
- They provide a coherent, complete set of variables describing the atmospheric state;
- They provide a reconstruction of the record of past weather since it is constrained by observations.

The computational effort to produce these reanalyses is generally very big. It requires the collection processing of millions of irregularly distributed and different types observations, the integration of a comprehensive weather model, and the processing of the vast amount of output data. The first global reanalyses (ERA-15, NASA/DAO and NCEP/NCAR) in the mid 1990s were therefore limited in resolution and time period they covered. Regional reanalyses for Europe, with global reanalyses data used at its boundaries, allows for more detail (see *figure 2*) and are developed in the EURO4M and this UERRA EU-projects. You will find an overview of current atmospheric reanalyses activities at https://reanalyses.org/index.php/atmosphere.

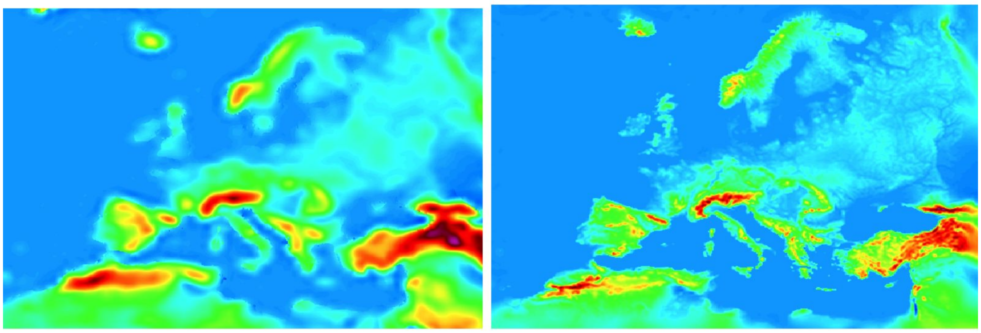

Figure 2: Global Reanalyses data from ERA-Interim (model resolution 80km) on the left, versus regional reanalyses data from EURO4M (model resolution 12km) on the right.

The UERRA reanalyses produced ensembles of regional reanalyses to enable an estimation of the uncertainties or biases in the datasets. Uncertainty in the reanalyses data may come from the observations, the model, the additional input (e.g. surface characteristics), as well as from the chaotic nature of the atmosphere itself. Apart from the estimate of the error in the basic meteorological reanalysis data, the ensemble approach also allows users to assess how this uncertainty cascades to the downstream applications, by using different ensemble members.

3. UERRA analyses systems and their products

This paragraph provides a short summary of the reanalysis systems used in the UERRA project and the output that is available to users.

The UK Met Office: UM 4D-Var and UM Ensemble 3D-Var

The EURO4M project (2010-2014) has provided the core 'deterministic' European regional reanalysis system, assimilating conventional, satellite and hydrological cycle (humidity, cloud, precipitation) observations into the Met Office Unified Model (UM)'s advanced four-dimensional variational (4D-Var, 12km resolution) data assimilation (Rawlins et al 2007). This system is run for the period 1979-2014
A coarser (36km) 3D-Var ensemble version is run for the same period with 20 ensemble members.
Both systems use ERA-Interim boundaries and cover a domain depicted in *Figure 3*. Data are made available through the ECMWF MARS archive.

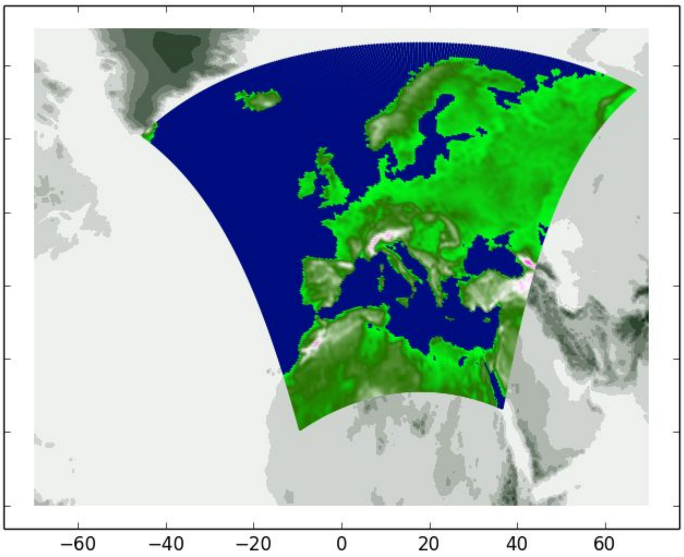

Figure 3: The domain of the UK Met Office deterministic and ensemble reanalysis products.

SMHI: 3D-VAR HARMONIE-ALADIN and HARMONIE-ALARO Ensemble

The HARMONIE Data Assimilation system is developed and used within the HIRLAM and ALADIN consortia is implemented and optimized for the entire European area with surrounding sea areas (see *Figure 4*) with a resolution of 11 km and 65 height levels. It is run for the period 1961-2015. For the period 1961-2001 ERA40 observations with addition of Swedish and French observations are used. After 2001 conventional data (SYNOP, Ship, Buoys, Radiosondes, Pilot and Aircraft) are used that are operationally available.
The system uses global reanalysis data as lateral boundaries: ERA40 for the 1961-1979 period, after that ERA-interim. Also, the large scales in the regional system are constrained by data from the global reanalysis.
An ensemble version of the system (HARMONY-ALARO Ensemble) is run for a shorter period: 2006-2015.
The results of both (deterministic and ensemble) systems are available through the ECMWF MARS archive.

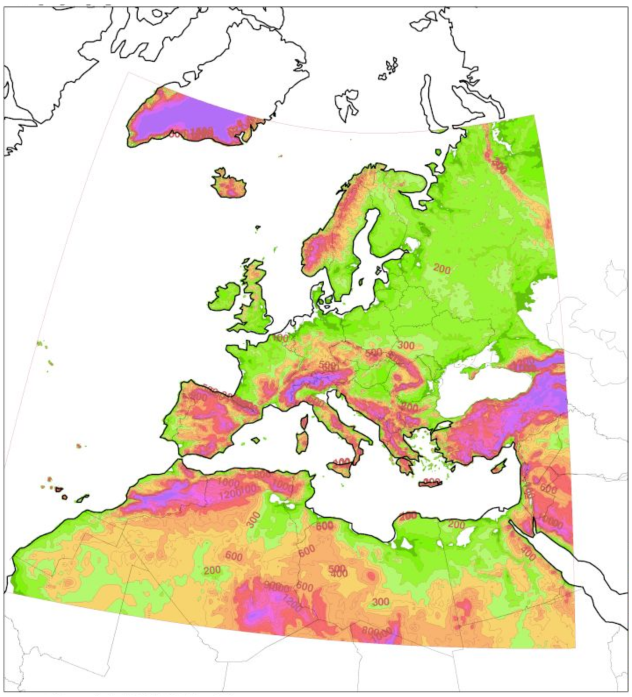

Figure 4: The domain of the SMHI HARMONY ALADIN/ALARO reanalyses products.

SMHI: MESAN Cloud V1 and Mesan Cloud V2 ensemble

Satellite observations of total cloud cover by the polar orbiting AVHRR, and the geostationary METEOSAT SEVIRI and MVIRI instruments are analyzed by SMHI to generate a high (11km) resolution cloud analyses for the period 2004-2008 (deterministic) and 1991-2010 (labeled ensemble but is only another realisation and will be produced later). When no satellite data are available, the cloud fraction from the SMHI HARMONIE reanalysis is used to fill the gaps. Access to the data is provided by the ECMWF MARS archive.

Météo France + SMHI: the MESCAN-SURFEX analysis and MESCAN-SURFEX Ensemble analyses

Météo France uses the 2D-analysis system MESCAN-SURFEX, to generate a surface analysis for temperature, relative humidity, precipitation, and wind. The system combines downscaled HARMONIE 3D-VAR reanalysis fields from SMHI and additional surface observation, to make a high resolution (5.5 km) 2-dimensional analysis over Europe for the period 1961-2015.
At the same resolution and domain an ensemble surface analysis consisting of 8 ensemble members (based on perturbed observations) is produced for the period 2006-2010. Both surface datasets are available through the ECMWF MARS archive.

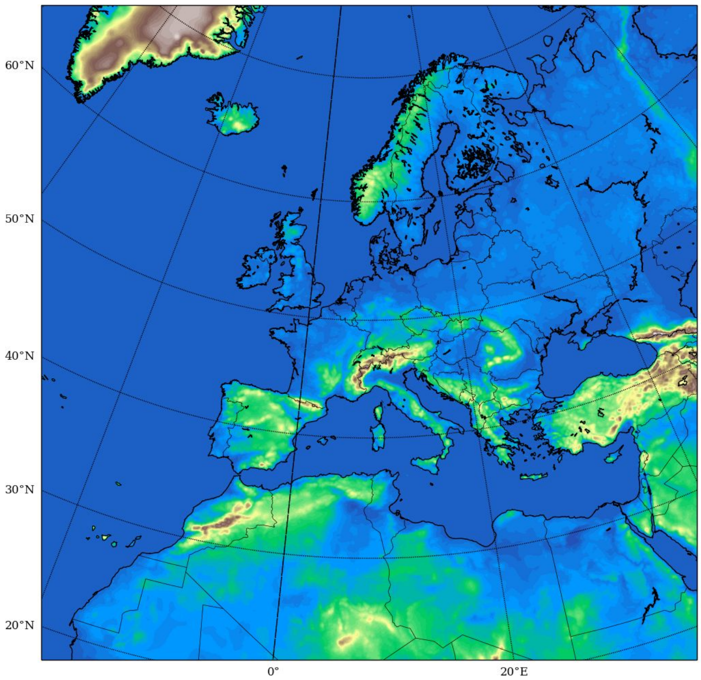

Figure 5: The Météo France MESCAN-SURFEX Domain

DWD and Univ. Bonn: COSMO Ensemble

An alternative to analysis/forecast cycles as depicted in *Figure 1*, is data assimilation that proceeds by some sort of continuous process such as *nudging*, where the model equations themselves are modified to add terms that continuously *push* the model towards the observations. The model used for this reanalysis is the COSMO EU model from the Deutscher Wetterdienst. An ensemble nudging method is used, where the 21 ensemble members are obtained by continuous relaxation of the model towards a set of perturbed observations (surface, radiosondes, and aircraft data). As boundary conditions the global reanalysis ERA-interim is used, as well as external analyses of sea surface temperature, soil moisture and snow. The period for which the reanalysis is produced is 2006-2010, with a spatial resolution of 12km over the domain shown in *Figure 6*. This dataset is available through the ECMWF MARS archive.

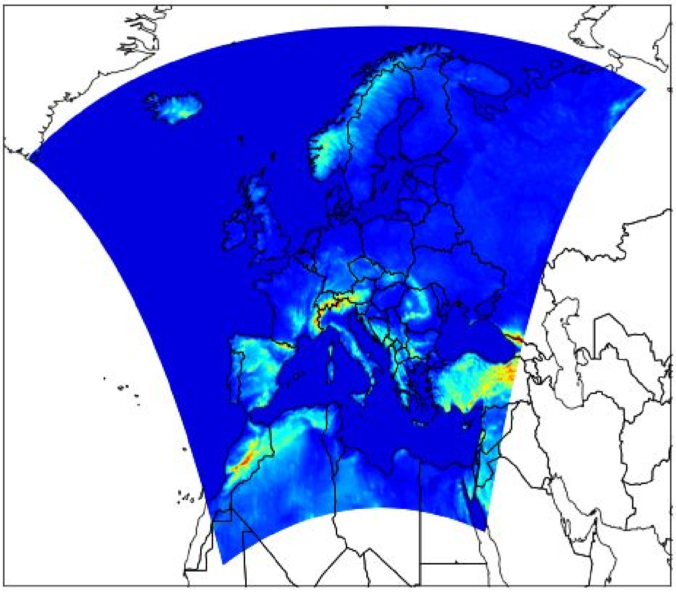

Figure 6: The COSMO-EU Domain

KNMI: E-OBS

The E-OBS gridded observation dataset is based on validated daily data from over 10,000 stations, and consists of fields of daily mean temperature, daily minimum temperature, daily maximum temperature, daily precipitation sum, and daily averaged sea level pressure. They cover the land surface in the area: 25N-75N x 40W-75E (see also *Figure 7*) and a period from 1950 up to present. The dataset is updated monthly , and full updates of the whole period are made every 6 months.

There are 4 different versions: 2 grid resolutions x 2 grid flavors. Data are made available on a 0.25 and 0.5 degree regular lat-lon grid, as well as on a 0.22 and 0.44-degree rotated pole grid, with the north pole at 39.25N, 162W.

The interpolation procedure used in E-OBS does not involve the use of a dynamical model, as in the other datasets of UERRA. The advanced statistical method provides 'best estimate' fields, and, in separate files, daily standard errors and elevation. See Haylock et al. (2008) and Van den Besselaar et al. (2011) for further details. Access to the dataset is provided at www.ecad.eu. The E-OBS data can also be accessed via the Copernicus Climate Change Service website (climate.copernicus.eu.).

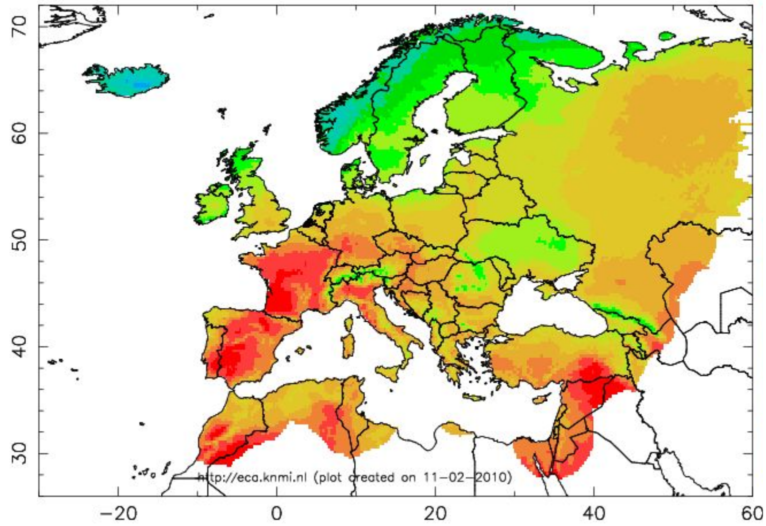

Figure 7: The E-OBS domain

Table 1: UERRA regional datasets that are made available to users.

Dataset	resolution	# of levels	period	type	Variables[1]
MetOffice UM	12km	70	1979-2014	4D-Var	Full set of variables
MetOffice UM	36km	70	1979-2014	3D-Var ensemble	Full set of variables
SMHI HARMONIE-ALADIN	11km	65	1961-2015	3D-Var	Full set of variables
SMHI HARMONIE-ALARO	11km	65	2006-2010	3D-Var ensemble	Full set of variables
SMHI MESAN cloud V1	11km	1	2004-2008	2D statistical interpolation	Total cloud cover
SMHI MESAN cloud V2	11km	1	1991-2010	2D statistical interpolation, Ensemble	Total cloud cover
Météo France + SMHI MESCAN-SURFEX	5.5km	1	1961-2015	2D statistical interpolation	T2m, RH2m, Precipitation
Météo France +SMHI MESCAN-SURFEX Ensemble	5.5km	1	2006-2010	2D statistical interpolation, Ensemble	T2m, RH2m, Precipitation
DWD COSMO-EU	12km	40	2006-2010	Nudging Ensemble	Full set of variables
KNMI E-OBS	0.25x0.25 (regular) – 0.22x0.22 (rotated pole)	1	2050-now	2D statistical interpolation	Daily surface Temperatures (Tmax, Tmin, Tmean), Precipitation and Pressure

[1]The full set of variables can be found in https://software.ecmwf.int/wiki/display/UER/Parameters

4. Applying and evaluating the data products

The regional reanalysis results are evaluated through comparison against independent, unassimilated observations. A basic intercomparison is made of the data from the different reanalysis datasets, the gridded observation datasets (e.g. the pan-European E-OBS dataset and the Alpine and Nordic gridded dataset, APGD and NGCD), and the global reanalyses datasets (e.g. ERA-CLIM).
In addition, a comparison was made of a limited set of application-oriented climate indices derived from the ensemble of reanalysis results.

The report on the evaluation of the datasets is available on the UERRA website http://uerra.eu/publications/deliverable-reports.html , please see D 3.6 (as the report will be finalised later than this text). In addition, a user-friendly summary will be made available soon on the UERRA website. These reports may not provide all the information a user may need to determine fitness-for-purpose. Additional evaluation of certain aspects of the reanalysis data in view of the specific requirements of the application may be needed. Some overarching questions will be relevant for all users, but the answers will depend on the application. These questions are, for instance:

Are reanalysis data accurate?
As an example, the evaluation revealed that the UERRA regional reanalyses systems tend to overestimate precipitation amounts and frequency, except COSMO. For some applications, e.g. in hydrology, it is therefore quite common to correct the precipitation data for a bias. Other variables, like surface temperature, are generally less variable in space and time and easier to reconstruct by the reanalysis system. Results in complex terrain, such as mountainous regions, or coastal areas, are generally less reliable than results over a more homogeneous terrain.
To make an assessment of the accuracy or to quantify the bias, a reference dataset (‘truth’) is needed for comparison. In comparisons against gridded station data, it should be realized that the quality of the data set is highly dependent on the station density which may vary from region to region. Satellite datasets may be used as a source of independent data for Europe, but the quality of, for instance, radiation observed by satellite is not sufficient to serve as a reference data set. Station data may be used as reference, since they are usually of high quality, but the coverage may be limited and there is the issue of representativity when point measurement are compared to grid cell values in the reanalysis dataset.
Finally, the fitness of the reanalyses results for a specific purpose also depends on the sensitivity of the application for errors in the input data.

Can we use extremes from reanalysis datasets?
Many applications have a focus on extremes, since they have the largest impact. In general, extremes are more difficult to reproduce than values closer to the mean. The value of precipitation extremes in the reanalyses data are highly dependent on the resolution of the dataset, and for gridded observations also on the station density. Extremes on larger scales, like droughts and heatwaves, may be represented much better. In general, representing absolute values of extremes is challenging for reanalysis systems. For instance, the distribution of areas with daily precipitation over a certain absolute threshold value is likely to be less accurate than the distribution using a relative threshold like a 95-percentile value.

Can we use reanalysis data for local applications?
Reanalysis data are gridded products. The values represent a certain spatial scale which may be hard to compare to a point value that may be obtained from a station. Note that the spatial scale of the data provided is not necessarily the grid spacing of the dataset. This is especially the case for the gridded ‘observation-only’ datasets like E-OBS and APGD, where areas with low station density are smoother due to statistical interpolation over large areas.
Downscaled datasets are useful when additional observations are added during the downscaling, or when

for instance more detail is added to the surface characteristics (e.g. orography). Examples of this are the MESCAN and MESAN datasets developed in the UERRA project.

Is the ensemble spread a good measure for the uncertainty?
Uncertainty in the reanalysis data may come from the observations, the model, the additional input (e.g. surface characteristics), as well as from the chaotic nature of the atmosphere itself. The evaluation shows that generally the single model ensembles represents only a part of the total uncertainty. In many cases the the single model ensembles show no or little overlap. The difference between the results of the reanalysis systems are larger than the spreads of the single model ensembles, indicating that the single model ensembles underestimate the uncertainty. However, it was also shown that the ensemble spread may be used as a qualitative indicator of the uncertainty. More details (different regions, variables, seasons) are provided in the 'Reanalyis Uncertainty Evaluation' which can be found on the UERRA website (Deliverable D2.14).

5. Data access and visualization

The UERRA reanalysis datasets (in total more than 700 Tb) are freely available at the MARS archive of ECMWF at https://apps.ecmwf.int/datasets/data/uerra. The full set of variables that are available can be found in https://software.ecmwf.int/wiki/display/UER/Parameters .
More details on how to retrieve the data from this archive, including some examples, conversion to NetCDF-format etc., are provided in the UERRA training material on the UERRA website (Deliverable 7.2).

For the full daily E-OBS dataset we refer to www.ecad.eu (for download and visualisation of E-OBS data).

An alternative is visualisation of E-OBS via the climate4impact portal using the ADAGUC package following this link: https://climate4impact.eu/impactportal/data/catalogbrowser.jsp?catalog=http://opendap.knmi.nl/knmi/thredds/./e-obs/e-obs-catalog.xml?

Currently a small subset of the data is also available through ESGF: https://esgf-data.dkrz.de/projects/esgf-dkrz under Project 'uerra'.

Details data access and visualization via ESGF and the climate4impact portal are described in the report 'Implementation of ESGF' on the UERRA website (Deliverable 4.4).