Seventh Framework Programme
Theme 6 [SPACE]

Project: 607193 UERRA

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

Deliverable D4.1
Subset of reanalyses in MARS at ECMWF

WP no:	4
WP leader:	Else van den Besselaar
Lead beneficiary for deliverable :	ECMWF
Name of author/contributors:	Richard Mladek, Manuel Fuentes
Nature:	Report
Dissemination level:	PU
Deliverable month:	48
Submission date: December 15, 2017	Version nr: 1

Report for Deliverable 4.1 (D4.1): Subset of reanalyses in MARS at ECMWF

By Richard Mladek and Manuel Fuentes

European Centre for Medium-range Weather Forecast

For this deliverable the goal is to archive agreed UERRA (Uncertainties in Ensembles of Regional Re-analysis) gridded reanalyses in ECMWF's archiving system MARS. To access and manipulate the stored data various tools were also prepared in the frame of the project.

Section 1: Building archiving tools

1.1 Introduction

In this section is described what had to be done before actual production archiving of UERRA reanalyses could start (the reanalyses are often referred to as datasets in the following).

A well designed data archive and tools to access and manipulate the archived data are crucial in projects like UERRA where huge amount of scientific data is produced and need to be easily accessible for users. ECMWF (European Centre for Medium-range Weather Forecast) has a lot of experience in these areas having been a partner in projects like TIGGE (The Observing System Research and Predictability Experiment 1), TIGGE-LAM (similar to TIGGE but for limited area weather prediction model outputs 2) or S2S (Sub-seasonal to Seasonal project 2). In all those projects, including UERRA now, the main task of ECMWF was to define data standards, data design and processing flows to do archiving work in the following step. Also related tools for data access and manipulation had to be created or updated for given new type of data.

1.2 UERRA website

A project website (related to data archiving part only) was established under wiki like system run at ECMWF (Error: Reference source not foundError: Reference source not found). This website has played a crucial role within the whole project and afterwards to keep all information in one place well organized and available to stakeholders. Below is an overview of the most important web pages there with short description. These pages will be maintained after the project end as well to assure a continuation of good service for users.

List of the main UERRA web pages at ECMWF:

- Description [4]
    - Parameters
        - list of archived parameters including their output frequency and all other details

    - Models
        - latest information about modelling systems which produced the final version of the archived UERRA data
- Support [5]
    - contacts, FAQ, User's forum
        - to keep the archive alive and updated after the project end
- Data production and archiving schedule [6]
    - which periods were archived
- Parameter availability [7]
    - which parameters are available for each dataset
- UERRA retrieval efficiency [8]
    - how to access data the best and most efficient way
- Issues with data [9]
    - list of acknowledged issues found in UERRA data

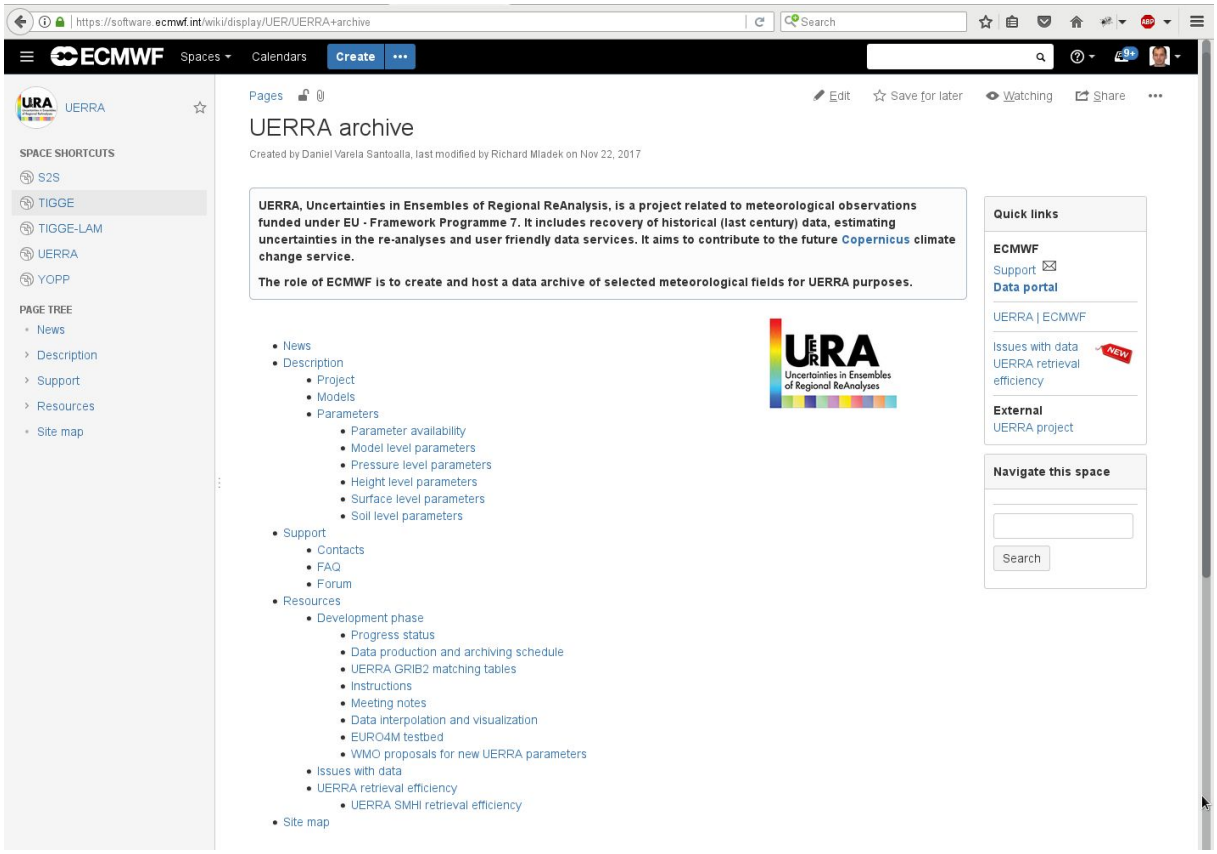

Figure 1 The UERRA home page at ECMWF

1.3 UERRA datasets

In Table 1 there is a final list of eight UERRA datasets which were produced and archived in MARS in the frame of the project. They are all regional reanalysis produced by four participating organizations on the defined UERRA domain covering mainly Europe and surrounding areas.

Model	**Organization**
COSMO/En	Meteorologisches Institut, Uni Bonn (MIUB)
HARMONIE/V1	Swedish Meteorological and Hydrological Institute (SMHI)
HARMONIE/V2	Swedish Meteorological and Hydrological Institute
MESAN	Swedish Meteorological and Hydrological Institute
MESCAN-SURFEX	Météo-France (MF)
MESCAN-SURFEX/En	Météo-France
UM	Met Office (MO)
UM/En	Met Office

Table 1 List of UERRA datasets

All scientific and technical details (which periods, parameters, levels, steps etc.) can be found in the above mentioned UERRA web pages at ECMWF.

There are two other UERRA datasets which might be still archived in MARS even after the project end (COSMO and MESAN/V2) depending on resources available on data providers' side (SMHI and DWD/MIUB). Technically MARS is ready to archive that additional data as it is very similar to the other listed and already archived datasets from those providers.

1.4 UERRA parameters

1.4.1 Description

The final list of UERRA parameters contains 38 surface (single level) parameters, 8 parameters on three types of vertical levels (model, pressure and height above surface) and 5 parameters on soil levels. The number of model levels varies among the models from 40 to 65. Similarly there are 3 to 14 soil levels in UERRA model outputs. Further there are two types of the archived model outputs forecasts and analysis with very complex and varying output frequencies (time steps) and models' start times (combinations of 0, 6, 12, 18 UTC) among the models. All relevant details can be again found in the Parameter web page [4] mentioned above. For illustration in Figure 2 there is a description of varying output frequencies for surface parameters.

1.4.2 Common definitions

Is it of utmost importance to have well defined data following the same standards in any successful archive. It must allow clear understanding and simple retrieval of each parameter without bigger additional effort on user's side. Failing to do so can lead easily to big and scientifically important but not too much used archives.

The GRIB2 format was chosen for archiving of UERRA data. It is the main archiving format for MARS and one of the most important meteorological data formats under WMO's (World Meteorological Organization) governance.

All parameters' definitions have to be further fully compliant with standardised WMO definitions which have been used for other similar datasets at ECMWF in the past (TIGGE, TIGGE-LAM, S2S). Thanks to that approach such data can be connected to many existing services using common standards for exchange of meteorological data and products.

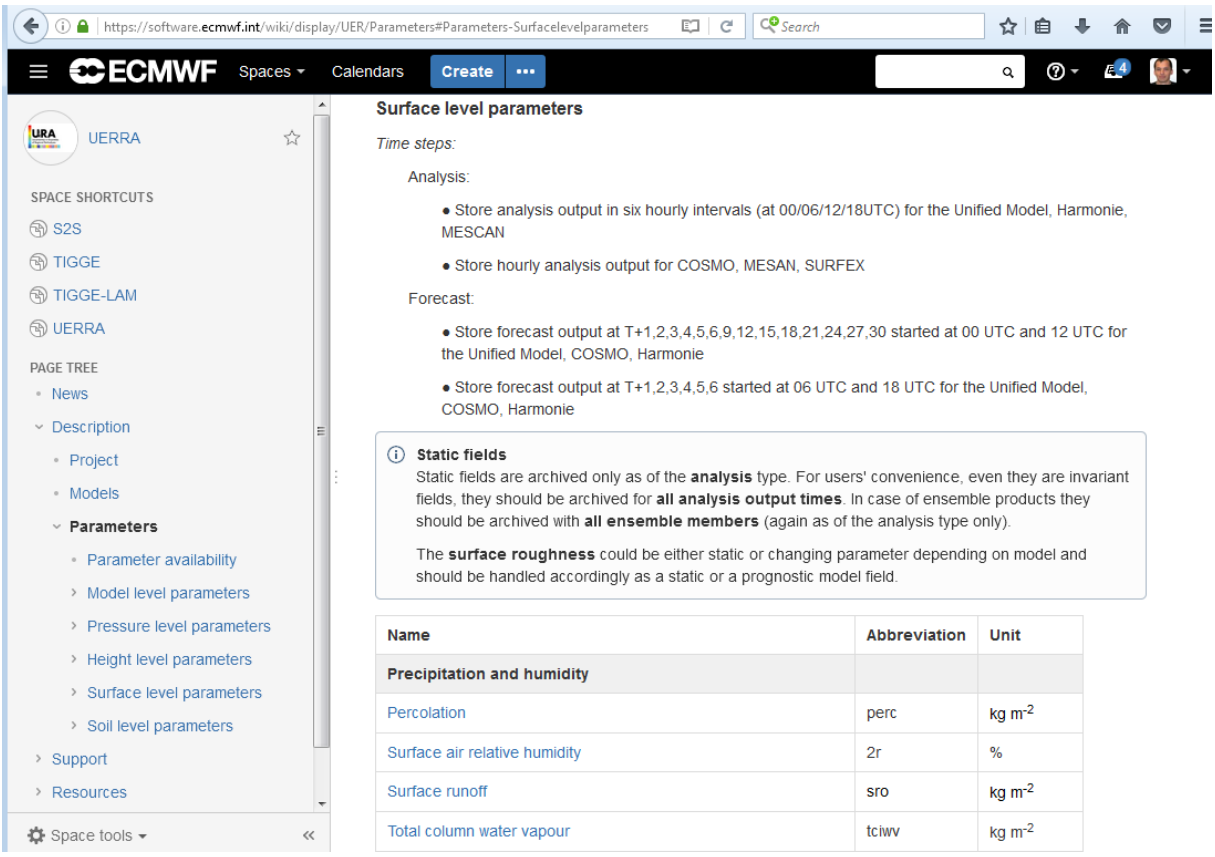

Figure 2 Output frequencies of UERRA surface parameters (screenshot of the web page Parameters at ECMWF)

Below is a short overview of the main achievements related to the work on common definitions of UERRA parameters in GRIB2 WMO compliant format.

- **Fully WMO compliant GRIB2 definitions were achieved for all UERRA parameters**
  - It took the first 3 years of the project to fully achieve this and it was one of the important "lessons learnt" for similar future activities. An example of GRIB2 encoding definition of one parameter is shown in Figure 3.

- **All UERRA parameters could be archived in MARS**

- The archive design has to be finished before any archiving work can start. Any data in MARS has to follow a certain structure based on archiving and retrieval patterns which must be know in advance. Again this step took almost first 3 years of the project.

- **New MARS developments for UERRA**
  - New level type *type=hl* (height levels above surface in metres)
  - New soil level type *type=sol*
    - Allowing varying grid dependant soil level depths
      - Not crucial for UERRA at the end as requested at the beginning but ready for coming future demands
  - 3 parameters “since previous post-processing” with varying validity periods

- **5 new UERRA parameters (WMO approval needed)**
  - Parameters below were not existing before the project start in WMO standard code tables and had to be proposed with all necessary details to be able to pass successfully the official approval process at WMO.
    - Percolation
    - Downward short-wave radiation flux, clear sky
    - Upward short-wave radiation flux, clear sky
    - Downward long-wave radiation flux, clear sky
    - Soil heat flux (updated version)

Althought some of the main achievements listed above may look rather like minor ones it is not true. It must be borne in mind the extent and complexity of MARS archive itself which belongs currently to the biggest meteorological archives in the world. Any even smallest change must be very carefully designed and tested before putting into production. It must not influence any other data already stored in MARS over almost 30 years of its existence. Because of the amount of parameters and all related metadata there such tasks have been never easy ones and work of various experts from ECMWF is always needed.

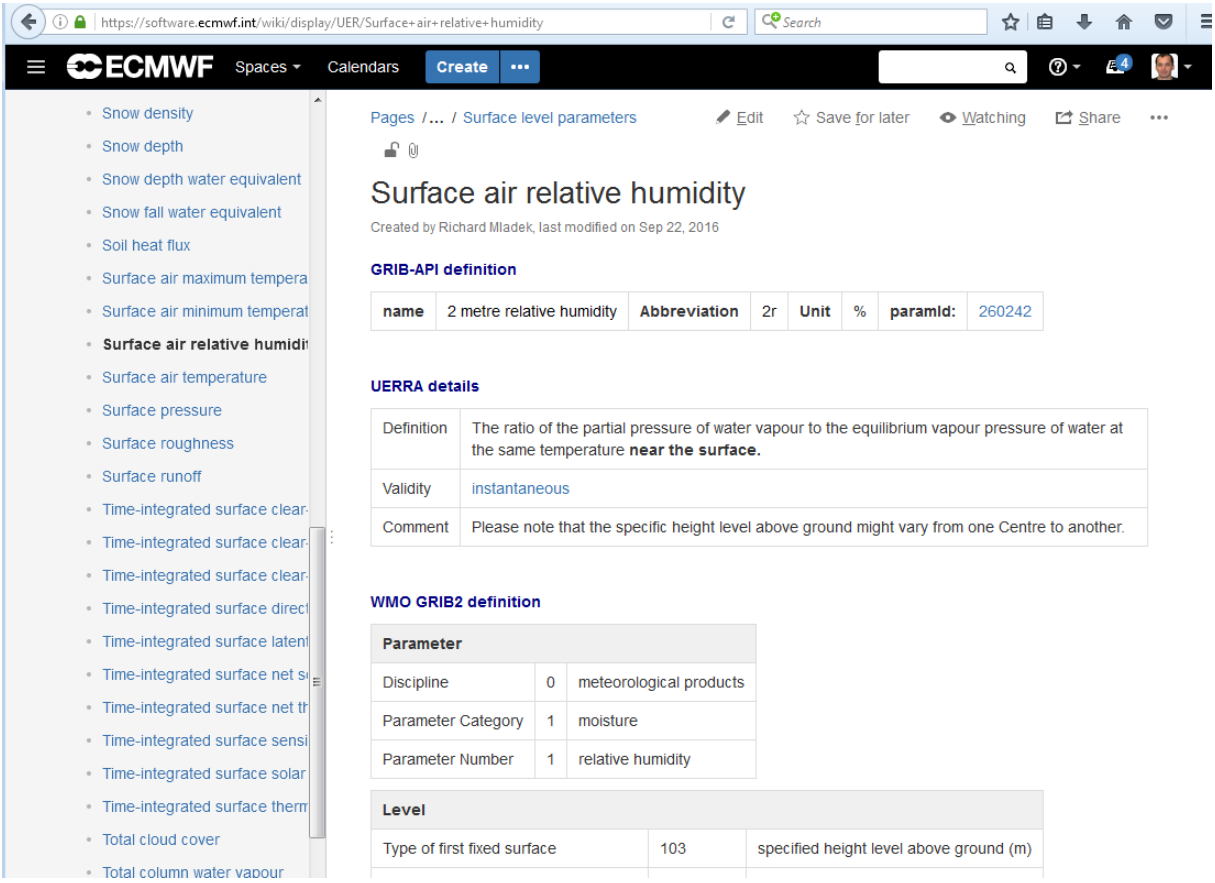

Figure 3 An example of GRIB2 encoding for surface air relative humidity (screenshot from UERRA web page at ECMWF)

1.5 Data handling tools

There are several tools necessary for smooth processing and maintenance of UERRA datasets. Those tools are based on already existing technologies used for other similar datasets mentioned above. Obviously many modifications were necessary to make those tools fully compatible with UERRA data. The work on the tools was done simultaneously most of the time, together with MARS design updates, as they depend on each other.

List of the data tools:

- **Data processing and archiving suite**
  - Collection of Shell / C / python codes for data processing (retrieval, encoding, modifications, verification etc.) run under ECFLOW (ECMWF task monitor scheduler)
- **Data portal**
  - Web interface to MARS for easy data access (see Section 2 for more details)
- **ECCODES [10]**

- Application program interface accessible from C, Fortran and Python programs developed for encoding and decoding WMO GRIB edition 1 and edition 2 messages

- **Web-API [11]**
  - Batch data access to data archived in MARS
- **METVIEW [12]**
  - Visualization and data handling software

The ECMWF Web-API, METVIEW and ECCODES are standard tools provided and maintained by the Centre.

Section 2: Archiving and accessing UERRA data in MARS

2.1 Archiving

The first test archiving attempts could start with the first UERRA data samples encoded to the defined WMO compliant GRIB2 files described in the previous section. That work started rather slowly in the first year of the project. Nevertheless the first fully production archiving of one of the UERRA datasets (HARMONIE/V1) could start only in the second half of the third year and was finished during the first half of the last project year 2017. Finally during that last UERRA year the production archiving of all other UERRA datasets could start as well. From the length of the period between receiving the first test samples to fully compliant ones ready for archiving one can guess how complicated the whole process was. It required a lot of work on both sides, ECMWF and providers', to fulfil all requirements. From that point of view the final technical work on archiving itself could be considered as rather easier one but still requiring a lot of time, careful checks and experience with production processing chains because of the data volume involved.

In Figure 4 is shown a progress status table related to UERRA datasets archiving as per December 2017. It can be seen that there was still ongoing archiving of three datasets with two of them expecting to finish only in 2018 i.e. already after the official project end.

	Model	Data status	Next milestone	By when	Progress log
1	COSMO (MIUB)	out of scope			UER-12 - C
2	COSMO/En (MIUB)	All data archived		Jul 2017	UER-12 - C
3	HARMONIE/V1 (SMHI)	All data archived		Jun 2017	UER-9 - HA
4	HARMONIE/V2 (SMHI)	All data archived		Aug 2017	UER-9 - HA
5	MESAN/V1 (SMHI)	All data archived		Oct 2017	UER-6 - ME
6	MESAN/V2 (SMHI)	out of scope			UER-6 - ME
7	MESCAN-SURFEX (MF)	Production archiving started	Complete production archiving	Xmas 2017	UER-10 - M
8	MESCAN-SURFEX/En (MF)	All data archived		Nov 2017	UER-10 - M
9	UM (MO)	Production archiving started	Complete production archiving	2018	UER-11 - U
10	UM/En (MO)	Production archiving started	Complete production archiving	2018	UER-11 - U

Figure 4 UERRA a progress status table (screenshot from UERRA web page at ECMWF)

The estimate of the final archive space occupied by UERRA data is around 800TB. At the beginning of December 2017, still during ongoing archiving, 700TB of UERRA data was already ingested by MARS. It represents more than 3.6 million fields.

In Table 2 there are listed final archived periods for each datasets. One of the biggest UERRA achievements is the length of those periods of homogeneously archived data compared to previous similar projects (EURO4M). It could be noted that the production, processing and archiving of Met Office reanalyses both deterministic and ensemble ones over the 37 years was probably one of the most expensive tasks ever run on ECMWF's computer premises and funded from allocated national quota for UK. The longest period archived is 55 years of deterministic outputs from HARMONIE/V1 model run by SMHI. The data volume per each provider as by December 2017 is listed in the last column of Table 2.

Model	**Period**	**Size [TB]**
COSMO/En	2006-2010	209
HARMONIE/V1	1961-2015	211
HARMONIE/V2	2006-2010	
MESAN	2006-2010	
MESCAN-SURFEX	1961-2000	85
MESCAN-SURFEX/En	2006-2010	
UM	1979-2016	218
UM/En	1979-2016	

Table 2 Final archived periods of UERRA datasets plus data volume per each provider.

2.2 MARS set up

In Table 3 there is shown the most important MARS set up relevant to UERRA data (MARS keys are used by the special MARS pseudo-meteorological language for easy access to the archive). Basic knowledge of that set up is needed for accessing UERRA data via Web-API or MARS catalogue (the latter only for users with full ECMWF account at ECMWF).

The top level MARS key is *class=ur* identifying UERRA data in the archive. There are further four *origins* used to identify the organization running the model, two *streams* to distinguish between deterministic and ensemble outputs (*oper* and *enda*) and three *types* for different model outputs (*an*, *fc*, *oi* for analyses, forecasts and optimal interpolation). It should be noted that HARMONIE/V2 was archived under ensemble *stream=enda* although it is a deterministic output. The reason is purely technical because of the too late full specification of that dataset when the MARS design could not be updated any more due to the already ongoing archiving that time. Similarly a new *type=oi* had to be invented and implemented in MARS to be able to archive MESAN dataset (*type=an* could have been also used if full details would be known earlier). Those small irregularities will not impact users and are made clear in the documentation to avoid potential confusion. It is another good example for "lessons learnt" how full data specification based on full sample from each expected dataset is important to have sufficiently enough in earlier stages of similar projects.

	MARS keys (class=ur, expver=prod)			
	origin	stream	type	number
COSMO/En	edzw	enda	an/fc	1..20
HARMONIE/V1	eswi	oper	an/fc	-
HARMONIE/V2	eswi	enda	an/fc	1
MESAN	eswi	oper	oi	-
MESCAN-SURFEX	lfpw	oper	an/fc	-
MESCAN-SURFEX/En	lfpw	enda	an/fc	1..8
UM	egrr	oper	an/fc	-
UM/En	egrr	enda	an/fc	1..20

Table 3 MARS keys set up for UERRA data

2.3 Accessing data

There are two main ways for users to access the UERRA data in MARS. The first is via a dedicated data portal which is intended for data discovery and shorter sample retrievals (up to one month). The second is so called Web-API interface which is a recommended way for downloading bigger data amount in a programmatic way via internet for use outside of the ECMWF. A screenshot of the dedicated UERRA data portal is in Error: Reference source not found. By clicking on required parameter a user can quickly check for which dates, times, steps etc. it is available. When the user selection is finished the data can be easily retrieved in GRIB2 format. Users often require to have data interpolated on different domain than the original one stored in MARS. Also sometimes NetCDF format is preferable to GRIB2. Generally MARS data portal has that kind of functionality for data interpolation or conversion to NetCDF. Unfortunately the data interpolation of limited area model

outputs like UERRA ones (with some specific domain geometries) are not fully supported yet. It should improve very soon with new interpolation package being tested. As for now an external package which could be used is documented in UERRA web page about interpolation and data visualization [13]. Also the conversion to NetCDF has been still only an experimental feature at ECMWF and because of that the conversion of UERRA GRIB2 to NetCDF is not fully working yet. Again there are many ongoing activities related to start of archiving NetCDF format in MARS and having fully working conversion from GRIB2 (which is at the moment still the main archiving format for MARS together with BUFR and ODB ones).

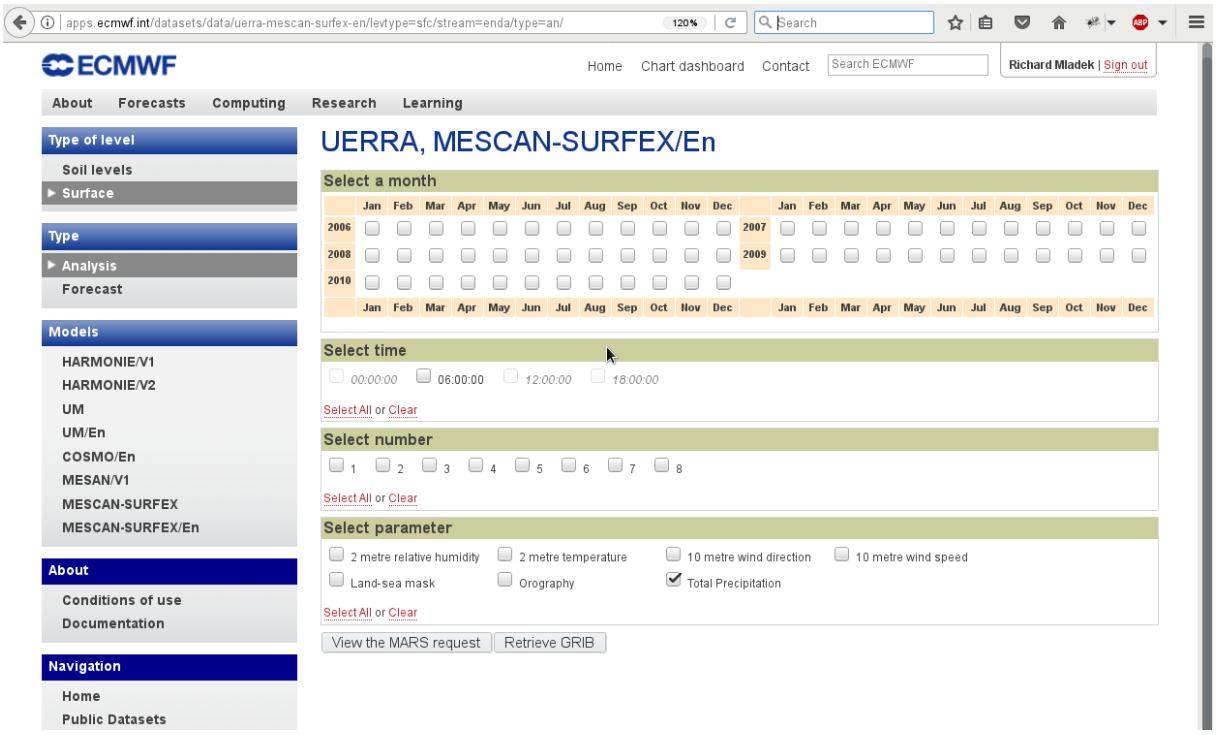

Figure 5 UERRA data portal. Example of access to MESCAN_SURFEX dataset for selected surface parameters from analyses.

References

1. https://software.ecmwf.int/wiki/display/TIGGE
2. https://software.ecmwf.int/wiki/display/TIGL
3. https://software.ecmwf.int/wiki/display/S2S
4. https://software.ecmwf.int/wiki/display/UER/Description
5. https://software.ecmwf.int/wiki/display/UER/Support
6. https://software.ecmwf.int/wiki/display/UER/Data+production+and+archiving+schedule
7. https://software.ecmwf.int/wiki/display/UER/Parameter+availability
8. https://software.ecmwf.int/wiki/display/UER/UERRA+retrieval+efficiency
9. https://software.ecmwf.int/wiki/display/UER/Issues+with+data
10. https://software.ecmwf.int/wiki/display/ECC
11. https://software.ecmwf.int/wiki/display/WEBAPI
12. https://software.ecmwf.int/wiki/display/METV/Metview
13. https://software.ecmwf.int/wiki/display/UER/Data+interpolation+and+visualization

Acronyms and abbreviations

COSMO	The Consortium for Small-scale Modeling
DWD	Deutscher Wetterdienst
ECMWF	European Centre for Medium-Range Weather Forecasts
EURO4M	The European Reanalysis and Observations For Monitoring
EPS	Ensemble Prediction System
GRIB	Gridded Binary
HARMONIE	**H**irlam **A**ladin **R**esearch on **M**eso-scale **O**perational **N**WP **i**n **E**uromed
MARS	Meteorological Archive and Retrieval System
MESAN	SMHIs Mesoscale Analysis System
MESCAN	“blending between MESAN and CANARI”
MF	Météo-France
MIUB	Meteorologisches Institut der Universität Bonn
MO	Met Office (UK)
NetCDF	Network Common Data Form
ODB	Observation Data Base
S2S	Sub-seasonal to Seasonal Prediction Project
SMHI	Swedish Meteorological and Hydrological Institute
THORPEX	The Observing System Research and Predictability Experiment
TIGGE	The International Grand Global Ensemble
TIGGE-LAM	TIGGE-Limited Area Model
UERRA	Uncertainties in Ensembles of Regional Re-analysis
WMO	World Meteorological Organization