Archiving of UERRA data in MARS and related data services

Richard Mládek, Manuel Fuentes,
Sebastien Villaume, Shahram Najm, Matthew Manoussakis

Highlights since the previous GA

+ Many full data samples got from providers
+ ECMWF's web pages with encoding and archiving instructions for UERRA completed
+ 1st UERRA production archiving has started (HARMONIE/V1)!

- Not all full data samples available yet
  - New problems might occur
  - Deterministic and ensemble outputs will be hopefully very similar (if not identical)
- Still unexpected flexibility needed to tackle UERRA data
  - Even now changing understanding of some UERRA parameters, levels etc.
  - Lack of the full samples 1-2 years ago

UERRA web pages at ECMWF

1. **Static pages**
   - New top level official project description
   - Parameters (with links to the encoding details for each parameter)
   - Instructions
     - Full UERRA gridded data test samples
     - GRIB2 encoding (including samples)
     - Archiving
2. **Tracking actual state**
   - Progress status (with links to the model specific log pages)
   - Parameter availability (based on the full sample data)
3. **UERRA data portal**
   - Production data only

UERRA web pages at ECMWF

*X~remark (explanation is below the table) **white**~not available **grey**~not applicable **green** ~available in analysis **blue**~available in forecast

	1 COS (edzw)		2 COS/En (edzw)		3 HAR/V1 (eswi)		4 HAR/V2 (eswi)		5 MESAN (eswi)		6 M-S (lfpw)		7 M-S/En (lfpw)		8 UM (egrr)		9 UM/En (egrr)		10 M-S (lfpw)	
	an	fc	an	fc	an	fc	an	fc	an	fc	an	fc	an	fc	an	fc	an	fc	an	fc
Model level (*3)																				
Cloud cover																				
Pressure																				
Specific cloud liquid water content																				
Specific cloud ice water content																				
Specific humidity																				
Temperature																				
U-velocity																				
V-velocity																				
Pressure level	*1	*1	*1	*1																
Cloud cover																				
Geopotential height																				
Specific cloud liquid water content																				
Specific cloud ice water content																				
Relative humidity																				
Temperature																				
U-velocity																				
V-velocity																				
Height level																				

GRIB decoder

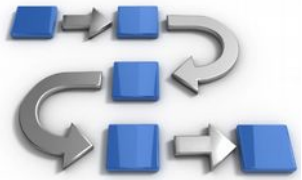
UERRA datasets - Progress status

	Model	Data status	Next milestone	By when	Progress log
1	COSMO (MIUB)	Full sample data provided	Start test archiving	Nov 2016	UER-12 - C
2	COSMO/En (MIUB)	Test data expected	Get full one day test sample	Nov 2016	
3	HARMONIE/V1 (SMHI)	**Production archiving started**	Complete production archiving	Dec 2016	UER-9 - HA
4	HARMONIE/V2 (SMHI)	Test data expected	Get full one day test sample	Dec 2016	
5	MESAN (SMHI)	Full sample data provided	Start test archiving	Nov 2016	UER-6 - ME
6	MESCAN-SURFEX (MF)	Test data received	Get full one day test sample	Nov 2016	UER-10 - M
7	MESCAN-SURFEX/En (MF)	Test data expected	Get full one day test sample	Nov 2016	
8	UM/4DVAR (MO)	Get full one day test sample	Start test archiving	Nov 2016	UER-11 - UI
9	UM/En4DVAR (MO)	Full sample data provided	Start test archiving	Nov 2016	UER-11 - UI

https://software.ecmwf.int/wiki/display/UER/Progress+status

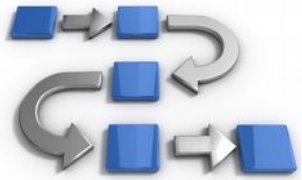
- The former MESCAN(MF) dataset renamed & split to 2 datasets (det & eps)
- 3 temporary datasets (short period; expver=test in MARS)
  - MESCAN-SURFEX **already archived** (less parameters; fc only +6H; 2008)
  - UM & UM/En (full data but not the final model version; 2008)

Data processing workflow



- Extract UERRA variables
  - Nothing should be missing or redundant
  - Double check UERRA final lists of parameters, levels, steps, etc.
- Encode/convert to UERRA-compliant GRIB2 format
- Run the checking tool
  - Use the latest tigge_check version (not a part of the official GRIB-API yet!)
- Transfer to ECMWF computers (if not already there)
- Archive in MARS (next slide)
  - By providing the exact archiving request
  - Check what was archived (no more fields, no less fields, no missing dates)

Archiving workflow



1. **MARSSCRATCH (expver=test/prod)**
   - Temporary development version of MARS
   - Anybody can use it for preliminary tests
   - Data can disappear at any time (not happening most of the time)
2. **MARS (expver=test)**
   - Final test before the full production archiving (samples should be already complete as in the next step)
   - Data might be overwritten
   - Used for some temporary (partial) UERRA datasets
3. **MARS (expver=prod)**
   - The final production version of the data which **might not** be overwritten
     - Any changes would impact users who already used the data!
     - Difficult to delete any archived data (e.g. by mistake)
   - One user ID per provider allowed only (but can be changed as needed)

UERRA datasets (MARS set up)

		MARS keys (class=ur)				
		origin	stream	type	number	remarks
Production datasets (expver=prod)						
1	**COSMO**	edzw	oper	an/fc	-	
2	**COSMO/En**	edzw	enda	an/fc	1..20	No full sample yet
3	**HARMONIE/V1**	eswi	oper	an/fc	-	**2008 archived**
4	**HARMONIE/V2**	eswi	enda	an/fc	1	2006-2010; not real eps; No full s. yet
5	**MESAN**	eswi	oper	TBD	-	**Data ready;** 1 param only (tcc)
6	**MESCAN-SURFEX**	lfpw	oper	an/fc	-	No full sample yet
7	**MESCAN-SURFEX/En**	lfpw	enda	an/fc	1..8	No full sample yet
8	**UM/4DVAR**	egrr	oper	an/fc	-	No full sample yet
9	**UM/En4DVAR**	egrr	enda	an/fc	1..20	
Temporary datasets (expver=test)						
10	**MESCAN-SURFEX**	lfpw	oper	fc	-	**2008 archived**; sub-set only
11	**UM/4DVAR**	egrr	oper	an/fc	-	2008
12	**UM/En4DVAR**	egrr	enda	an/fc	1..20	2008

General problems

- Still continuous new information coming from the recent data samples affecting production archiving start (agreement & decisions needed; minor GRIB-API & MARS design for UERRA updates)
  - Soil levels/layers (varying depths x fixed; semi-levels)
  - Model levels/layers (semi-levels)
  - Parameters since the previous post-processing (exact meaning)
  - Static fields handling (how often to archive)
- Data checking tool (tigge_check) updates and tuning
  - Must reflect recent findings from the full sample data
  - Instant allowed limits tuning for UERRA because of its specifics (domain size, 1H outputs, high altitudes => e.g. albedo or cloud cover might be 0)
- Interpolation of rotated lat-lon not working yet but new ECMWF's MIR (Meteorological Interpolation and Regridding) tool in preparation
- Conversion GRIB2 to NetCDF of UERRA data not working yet

Model specific features

- **COSMO**
  - Model levels above 100 hPa not available
  - Step 0 not available in forecasts ("fc") because of nudging
- **MESCAN-SURFEX**
  - Contains analysis of total precipitation accumulated between 6H of the previous day and 6H of the day encoded in the GRIB2
  - Background fields produced (MARS not ready)
- **HARMONIE**
  - Only steps up to +6H are available for soil level parameters
- **UM**
  - Model level increases with height MO (opposite to other models)

Model specific features

- **Model levels**
  - **COSMO:** 1..40
  - **HARMONIE:** 1..65
  - **UM:** 1..63
- **Soil levels**
  - **COSMO:** 8 soil layers
    level borders at 0, 0.01, 0.03, 0.09, 0.27, 0.81, 2.43, 7.29, 21.87 m
  - **HARMONIE:** 3 soil levels/layers (sot on levels, vsw on layers)
    level depths are grid dependent
  - **MESCAN-SURFEX:** 14 soil layers (only 6 layers for preliminary shorter re-analysis runs)
    level borders at 0, 0.01, 0.04, 0.1, 0.2, 0.4, 0.6 m
  - **UM:** 4 soil layers
    level borders at 0, 0.1, 0.35, 1 and 3 m

HIRLAM datasets

- 1st UERRA production archiving (HARMONIE/V1)!
  - Scripts for UERRA parameters' selection, processing, conversion and archiving provided
    - Further modified and used by Esbjörn Olsson
    - Many sample data exchanged before reaching the final state
    - Some suspicious fields removed in the last minute before archiving start (percolation, surface runoff)
    - Varying soil depths are not specified yet (details needed)
  - 2006-2008 completed
    - Grand total: 3,527,579,450,865 (3.20831 TiB)
    - Number of fields: 5,052,264
    - Bug found and fixed (two parameters parameter number swapped)
  - Double archiving in progress (2 year in parallel)

HIRLAM datasets

- Originally HARMONIE/V1 and HARMONIE/V2 should have been archived under stream=enda like ensemble system with members 1 and 2 (not the best solution either as it is not a real ensemble)
  - At the end HARMONIE/V1 (the main product) has been archived under stream=oper
  - HARMONIE/V2 will be archived under stream=enda and number=1
  - Problem now how to archive the 3rd HIRLAM dataset MESAN (new type request submitted to ECMWF's Data Governance Group)

- These problems could have been avoided if full samples were available before MARS design

COSMO datasets

- Excellent cooperation with COSMO team (Lilo & Dörte Liermann & Ulrich Schättler)
  - Should be in MARS soon!

- Scripts for UERRA parameters' selection, processing, conversion and archiving provided; now being modified and tested by COSMO team
  - u/v conversion to wind dir/speed on hl levels
  - Conversion of volumetric soil parameters' units from m3/m3 to kg/m3 and required soil layers encoding
  - Adding bitmap to soil fields to mask correctly undefined values
  - Orography (Z0) rescaling have to have correct units
  - Constant fields replication to be able to archive them with each "an" time
  - Additional post-processing of hl level data needed (from height above sea level to height above surface)

COSMO specifics

- **Model level data**
  - Non-hydrostatic model (layers, semi-levels)
  - Will be archived on full levels (details in documentation)
  - Vertical coordinate parameters PV for the half levels
  - GRIB2 level type 118 instead of 105 (next release of GRIB-API v.1.19.0)
- **Soil level data**
  - Will be archived on full levels (details in documentation)
- **COSMO local GRIB2 tables**
  - Maintained as a part of ECMWF's GRIB-API
  - Problems with conversion to GRIB2 (workaround needed)
    - Snow depth water equivalent
    - Snow fall water equivalent

Met Office datasets

- Data already in GRIB2
- The full samples have been in good shape since the beginning (Peter Jermey)
  - Still to achieve fully UERRA compliant format takes time
  - MARS archiving scripts and data checking tools provided
  - UM/En4DVAR data technically ready for MARS
- Full production runs not finished yet (some observation types and hybrid DA system still need work)
  - Agreed to archive firstly the available data under expver=test in MARS for 2008 only (both streams oper and enda) by the end of 2016
  - The final version of the data will be archived next year (the same content but for the full UERRA period)

Météo-France datasets

- MESCAN-SURFEX archived in MARS (by Rachid mainly for WP3 needs)
  - Temporary sub-sample from 2008 under expver=test
    - 5 parameters from reanalyses (2t, 10wdir/ws, 2r, tp)
    - 13 parameters from reforecasts (+6H only)

- The full data not ready yet
  - but MESCAN-SURFEX partial archiving could start
    - Total precipitation (6 eps members)
    - 2t, 2r (8 eps members)
    - As it is the final version it can be archived under expver=prod already and the rest will be added afterwards once available

ECMWF links for UERRA

- UERRA at ECMWF (high level page): http://www.ecmwf.int/en/research/projects/uerra
- UERRA at ECMWF: https://software.ecmwf.int/wiki/display/UER
- EURO4M testbed data in MARS: https://software.ecmwf.int/wiki/display/UER/EURO4M+data+in+MARS
- Progress status: https://software.ecmwf.int/wiki/display/UER/Progress+status
- Parameter list: https://software.ecmwf.int/wiki/display/UER/Parameters
- Parameter availability: https://software.ecmwf.int/wiki/display/UER/Parameter+availability
- Instructions: https://software.ecmwf.int/wiki/display/UER/Instructions
- Data interpolation and visualization: https://software.ecmwf.int/wiki/display/UER/Data+interpolation+and+visualization