

## **Wikiprint Book**

**Title: DATASETS**

**Subject: TracMeteo - udg/ecomms/dataserver/catalog**

**Version: 116**

**Date: 05/16/2022 01:57:59 PM**

## Table of Contents

<b>DATASETS</b>	<b>4</b>
WFDEI (provided by the EU-funded WATCH project)	4
?NCEP_ reanalysis1 (provided by NCEP/NCAR)	4
ERA_interim (provided by ECMWF)	4
System4 (provided by ECMWF)	4
CFSv2 (provided by NCEP)	4
GloSea5 (provided by Met-Office)	4
GloSea5 (12 members)	5
GloSea5 (24 members)	5
SMHI-EC-EARTH_EUPORIAS hindcast (provided by SMHI)	5

The **ECOMS UDG** collects and provides information (mainly at 6-hourly and/or daily resolution) for a reduced number of variables from a number of datasets (seasonal hindcasts, reanalysis and observations) obtained from different data providers. The following list of variables has been identified according to the user's needs, receiving feedback from EUPORIAS WP22 (climate information indices, CIIs), WP23 (impact models), WP21 (calibration and downscaling) and SPECS WP61 (pilot applications) and WP52 (calibration and downscaling). See the section on the [assessment of user's needs](#) for more details.

Note that the **R names** below correspond to the vocabulary names used in the [R data access package](#) for homogenization purposes. Note that, data homogenization and aggregation (i.e. daily means from 6h data) is only provided through the R data access package.

In order to specify the particular **temporal frequency/aggregation** available for the variables in the different datasets, the following codes are used in the table below: **6h** (6-hourly instantaneous data). **12h** (12-hourly instantaneous data). **24h** (24-hourly instantaneous data). **DM** (daily mean value). **DX** (daily maximum value). **DN** (daily minimum value). **DA** (daily accumulated data). **DAR** (accumulated since the initialization time `?runtime`) NOTE: The R package performs deaccumulation on a daily basis to match the standard definition. **fx** (static field)

In the table below, boldface codes (e.g. **6h**) indicate variables already available through the ECOMS UDG. **Italics** are used for work in progress (variables to be included in the next update). **e** indicates that a variable exists in the original dataset but it is not planned to be included yet in ECOMS-UDG; **blanks** indicate that the variables do not exist in the original dataset. Codes ended by (\*) indicate variables which do NOT exist in the dataset, but are derived/approximated from other available ones through the [R data access package](#). More information on the particular approximations used are given in the [?variables-datasets mapping](#). Variables ended by (#) indicate daily aggregated values obtained from the corresponding original 3-hourly data.

R name	Variable description	Observations:			Reanalysis:		Seasonal forecasting models:					
		WFDEJ	NCEP_reanalysis1	ERA_interim	System_seasonal_18	System_seasonal_21	System_seasonal_18	CFR2_seasonal_18	Seasonal_18	Seasonal_21	Seasonal_18	
<b>Surface variables</b>												
tas	Near-Surface air temperature	DM	6h	DM	6hDM	DM		e			DM	DM
tasmax	Daily Maximum Near-Surface Air Temperature	DX(#)	6h	DX	DX	DX	DX	6h			DX	DX
tasmin	Daily Minimum Near-Surface Air Temperature	DN(#)	6h	DN	DN	DN	DN	6h			DN	DN
tp	Total precipitation amount	DA	6hA	DA	DAR	DAR	DAR	6h			DA	DA
psl	Sea Level Pressure		6h	DM	6h	6h	12h	6h			DM	DM
ps	Surface air pressure	DM		e	6h(*)			e				
wss	Wind speed (at 10m)	DM		e	6h(*)	e	e	6h(*)				
tdps	2m Dewpoint Temperature			e	6h	e						
hus	Surface (2m) specific humidity	DM	6h	e	6h(*)			e				
rsds	Surface Downwelling Shortwave Radiation	DA	6hA	e	DAR	e		e			DA	DA
rls	Net Longwave Surface Radiation	DA	6hA	e	DAR	e		e				
sst	Sea surface temperature			e	e	e	e					
uas	Eastward Near-Surface Wind		6h	e	6h	e	e	6h				
vas	Northward Near-Surface Wind		6h	e	6h	e	e	6h				
wssmax	Wind speed (at 10m)			e	e	e	e	e				
wgust	Wind gust			e	e	e						
mrs0	Total Soil Moisture Content							e				
mros	Surface runoff flux				e	e		e				
mr0	Total Runoff				e	e						
sr0	Sub-surface runoff rate				e	e						
wcsl	Water Content of Soil Layer				e	e						
psn	Snowfall amount	DA			e	e						
sd	Snow Depth				24h	e		e				
<b>3D vars @ isobaric surface levels</b>												
ua	Eastward Wind		6h @ standard levels	DM @ standard levels	12h @ standard levels	e	e	e			6h @ standard levels	
va	Northward Wind		6h @ standard levels	DM @ standard levels	12h @ standard levels	e	e	e			6h @ standard levels	
zg	Geopotential height		6h @ standard levels	DM @ standard levels	12h @ standard levels	e	e	e			6h @ standard levels	
ta	Air temperature		6h @ standard levels	DM @ standard levels	12h @ standard levels	e	e	e			6h @ standard levels	
hus	Specific humidity		6h @ standard levels	DM @ standard levels	12h @ standard levels	e	e	e			6h @ standard levels	
<b>Static fields</b>												
zgs	Surface geopotential height				fx	e		e				

@ standard Levels: 1000,850,700,500,300,200 mb, except for hus, which is not available at 200mb in some models

**Data Homogenization:** The different nature of the datasets, and the idiosyncratic naming and storage conventions often applied by the modelling centres, makes necessary an homogenization across datasets in order to implement a truly user-friendly toolbox for data access. To this aim, the [R package for data access](#) has been developed. Data homogenization is achieved through the creation of a common vocabulary. The particular variables of each dataset are then translated -and transformed if necessary- into the common vocabulary by means of a *dictionary*. Both features -vocabulary and dictionary- are described [here](#). In particular, some typical transformations performed by the `loadECOMS` interface are deaccumulation of initialization-accumulated variables to daily accumulated (i.e.: **DAr** --> **DA**) and scaling and/or offset of variables to match standard units (e.g. -273.15 for conversion K --> °C).

## DATASETS

### WFDEI (provided by the EU-funded WATCH project)

The WATCH-Forcing-Data-ERA-Interim: [?WFDEI](#) was produced post-WATCH using WFD methodology applied to ERA-Interim data. It is a meteorological forcing dataset extending into early 21st C (1979 ? 2012). Eight meteorological variables at 3-hourly time steps, and as daily averages, for the global land surface at 0.5° x 0.5° resolution.

### ?NCEP\_reanalysis1 (provided by NCEP/NCAR)

A subset of predictors commonly used in statistical downscaling.

### ERA\_interim (provided by ECMWF)

A subset of predictors commonly used in statistical downscaling (daily means at 2° resolution). This information has been downloaded from the ECMWF's MARS, degraded to a common 2 degrees grid and post-processed computing daily means based on the original 6 hourly fields when required. Therefore, this dataset is a degraded and subset of the original ERA-Interim reanalysis dataset, which is freely available via [?ECMWF servers](#) at original resolution.

This dataset is used in the framework of different international initiatives (such as CORDEX-ESD and VALUE) in order to have a standard predictor dataset and to facilitate the work of the contributing downscaling groups. More information [?here](#).

### System4 (provided by ECMWF)

The [?System 4](#) seasonal forecasting system became operational in November 2011. The corresponding hindcast is archived in the Meteorological Archival and Retrieval System ([?MARS](#)), the main data repository at the ECMWF, as a collection of GRIB-1 files at 0.75° spatial resolution. The downloaded data has been exposed as three different virtual datasets:

- [?System4\\_seasonal\\_15](#): There are twelve initializations (hereafter called *runtimes*) per year (the first of January, February, ...), each with 15 members running for 7 months (hereafter called simply *times*). Period: 1981-2010.
- [?System4\\_seasonal\\_51](#): There are only four *runtimes* per year (the first of February, May, August and November), each with 51 members running for 7 months. Period: 1981-2010.
- [?System4\\_annual\\_15](#): There are four *runtimes* per year each with 15 members, but the forecasts run for 13 months. Period: 1981-2010.

A preliminary [validation report](#) produced in SPECS (milestone MS22) is available for precipitation (`System4_seasonal_15`). The reports for all datasets and variables will be produced after feedback with end-users.

### CFSv2 (provided by NCEP)

The [?CFS version 2](#) seasonal forecasting model became operational at NCEP in March 2011. The corresponding [?retrospective CFSv2](#) forecast dataset is stored in the [?NOMADS server](#) as a collection of GRIB-2 files at 1° spatial resolution. The downloaded data is exposed as a single virtual dataset:

- [?CFSv2\\_seasonal](#). There are four initializations (4 cycles) from every 5th day (thus providing on average 24 members per month) running for 9 months (see [?the CFSv2 member configuration](#) for more detailed information of member construction). Period: 1982-2010. **Note:** For better comparability with other hindcasts, the [R data access package](#) defines by default an ensemble of 15 members for each lead month and forecast season.

### GloSea5 (provided by Met-Office)

GloSea5 data have been exposed as two different virtual datasets through ECOMS-UDG, based on their different forecast extents, initializations and member configuration. Members are defined from all combinations of the members and lagged initializations.

In particular, in UDG members are ordered considering all combinations of member  $\times$  runtime, being member the factor that varies fastest. For instance, a call to `loadECOMS` using `members = 1:5` in `dataset = Glosea5_seasonal_12` will return, in this order, the following 5 members:

- Member1: [member1, runtime1]
- Member2: [member2, runtime1]
- Member3: [member3, runtime1]
- Member4: [member4, runtime1]
- Member5: [member1, runtime2]

Their main characteristics are next briefly enumerated:

#### **GloSea5 (12 members)**

- 4 initializations per year: Feb, May, Aug and Nov
- $4 \times 3$  members per initialization
- Total of 168 runtimes and 130 forecast times per runtime (~4.5 months)
- Hindcast period: 1996-2010

#### **GloSea5 (24 members)**

- 2 initializations per year: May and Nov
- $8 \times 3$  members per initialization
- Total of 126 runtimes and 120 forecast times per runtime (~4 months)
- Hindcast period: 1992-2013

#### **SMHI-EC-EARTH\_EUPORIAS hindcast (provided by SMHI)**

The downloaded data is exposed as a single virtual dataset:

- [?SMHI-EC-EARTH\\_EUPORIAS](#) hindcast. There is one initialization, corresponding to May and 15 members running for 4 months. Period: 1991-2012. **Note:** For better comparability with other hindcasts, the [R data access package](#) defines by default an ensemble of 15 members for each lead month and forecast season.

See the [UDG wiki](#) for additional groups and datasets provided by the UDG.