# ERA-Interim Daily Climatology

*Martin Janoušek, ECMWF*
*January 2011*

The daily climatology represent a derived product of a re-analysis daily analysis dataset. It provides a best estimate of climate characteristics for the given day of the year in the form of  mean, standard deviation, and other statistics describing climate distribution in grid points of a global grid.

The primary application of the daily climatology is the evaluation of anomalies of atmospheric parameters, both in forecast fields (e.g. for anomaly probability prediction by EPS) and in analyses. A good daily climatology is essential for computing of anomaly-based verification scores, both deterministic (anomaly correlation coefficient, mean anomaly of forecast etc.) and probabilistic (Brier score, ranked probability score, skill scores etc.). The presence of detailed distribution characteristics of the daily climatology makes it also suitable to serve as a reference forecast method for skill scores calculations.

At ECMWF the methodology for construction of the daily climatology has been developed and tested for ERA-40 dataset (Jung and Leutbecher, 2008) on the period 1979-2001. Currently the new daily climatology based on 1989-2008 analyses of ERA-Interim has been built well. It is used by the new operational verification system at ECMWF and it is now available for other users' applications.

This memorandum briefly summarises the method of daily climatology construction and lists available parameters and levels.

## Product overview

ERA Interim daily climatology (ERAI-DACL) consists of horizontal fields of statistical characteristics of atmospheric parameters at a given time and day of a year. For a given parameter and level and a given day of a year and time there are fields of

- mean (a field of the filtered mean values of the parameter valid at a given hour and on a day of a year computed from the sampling 20-years period),
- standard deviation (idem for standard deviation),
- minima and maxima (idem for extreme values)
- quantiles (the terciles and selected percentiles of the distribution of the parameter during the sampling period).

The ERAI-DACL fields are available from ECMWF MARS in the GRIB-1 format. The upper-air parameters are available at 00 and 12 UTC network times and the surface and screen-level parameters at times 00, 06, 12, 18 UTC. Due to their high computational cost and data size the quantiles are available only for selected parameters. Climatology for wave parameters is available as well using different STREAM DACW.

### *The list of parameters and MARS identifiers*

- **class**=EI
- **stream**=DACL (or DACW for wave parameters swh and mwp)
- **expver**=0001
- **type**=EM*(mean)*/ES*(standard deviation)*/FCMIN*(minima)*/FCMAX*(maxima)*
    - for **time**=00:00/12:00
        - **levtype**=PL
          **levelist=**1/2/3/5/7/10/20/30/50/70/100/150/200/250/300/400/500/600/700/850/925/1000
          **param**=Z/T/U/V/WIND/W/R/Q/O3

        - **levtype**=PT
          **levelist**=300/315/330/350/370/395
          **param**=PV
        - **levtype**=PV
          **levelist**=2000
          **param**=PT
    - for **time**=00:00/06:00/12:00/18:00
        - **levtype**=SFC
          **param**=2T/10U/10V/10SI/MSL/SSTK/CI/STL1/ISTL1/SWH/MWP
- **type**=PB*(quantiles)*
  **quantile**=1:100/2:100/5:100/10:100/15:100/20:100/25:100/30:100/1:3/35:100/40:100/45:100/50:100/55:100/60:100/65:100/2:3/70:100/75:100/80:100/85:100/90:100/95:100/98:100/99:100
    - At **time** 00:00/12:00:
        - **levtype**=PL
            - **levelist**=50/100/200/250/500/700/850/1000, **param**=Z/T
            - **levelist**=200/250/500/700/850, **param**=U,V,Q
            - **levelist**=850, **param**=WIND
        - **levtype**=SFC, **param**=2T/10U/10V/10SI/MSL/SWH/MWP
- to specify **date** choose any leap year (e.g. 1994)

# Methodology for daily climatology

The methodology for construction of the daily climatology has been developed and tested on ERA-40 dataset (Jung and Leutbecher, 2008).

Here, we compute climatological statistics which depend on location and the day of the year. The climatology consists of daily fields of the mean, standard deviation of anomalies, and quantiles of anomalies. The climatology is based on ERA-Interim analyses, which are expected to provide the most accurate available, consistent, long-term description of the atmosphere. The climate is based on the 20 years 1989-2008. For each day of the year, statistics are based on a 61-day window centred on the day of interest. The statistics are computed with weights which are maximum at the window centre and gradually decrease to zero at ±30 days. Thus, 20×61 = 1220 dates contribute to the climate statistics of one day. Variable weights are superior to constant weights in terms of resolving the annual cycle and in filtering high-frequency sampling uncertainty. The climate of a particular wavenumber band is computed from the filtered analyses. For the climatological mean, one could alternatively compute the mean of the full fields and then apply the various filters. However, for the non-linear statistics, such as standard deviation and quantiles, the computation of the statistic and the waveband filter are not commutative.

The climatological statistics is computed from analyses for $N_Y = 20$ years. For day $\nu$ in month $\mu$, the climatology is computed from data $N_{1/2}$ days before the day to $N_{1/2}$ days after. The choice of $N_{1/2}$ has to seek a compromise between well resolving the annual cycle and a sufficient sample size. It was chosen $N_{1/2} = 30$. Due to the variable weights this fairly large window still resolves the annual cycle well. In Jung and Leutbecher, 2008, climatologies were also computed with $N_{1/2} = 10$ and 5. They still showed some noise on time scales smaller than one month but did not seem to contain additional information about the annual cycle.

The dates are arranged in a periodic manner in order to obtain a continuous annual cycle without a jump at New Year. It is convenient to express the dates in terms of Julian day number. Let $J_{YMD}(j_Y, j_M, j_D)$ denote the Julian day number for day $j_D$ in month $j_M$ and year $j_Y$. The climate is computed from data between 1 Jan 1989 and 31 Dec 2008. In order to formulate the periodicity, we define $J_0 = J_{YMD}(1989,1,1) = 2447528$ and

$\Delta J = J_{YMD}(2009,1,1)$ -$J_0$ = 7305. Any date $J$ is mapped to date

$$J' = J_0 + \{(J - J_0) \bmod (\Delta J)\} \tag{1}$$

which always falls between 1 Jan 1989 and 31 Dec 2008.

The climatology for day $\nu$ and month $\mu$ is computed from dates

$$D_{jk}(\mu,\nu) = \{J_{YMD}(1989+k,\mu,\nu)+j\}'$$

where index $j = -N_{1/2}, -N_{1/2}+1, \ldots, N_{1/2}$ specifies the distance in days from the centre of the data window and index $k = 1,2,\ldots,N_Y$ specifies the year. Thus, the total sample size is $N_Y(2N_{1/2}+1) = 1220$. The statistics are computed with a weighted average designed to damp high-frequency sampling uncertainties. The weights depend on the distance $j$ from the centre of the window

$$w_j = \frac{3(N_{1/2}+1)}{N_Y(2N_{1/2}+1)(2N_{1/2}+3)}\left\{1-\left(\frac{j}{N_{1/2}+1}\right)^2\right\} \tag{2}$$

Let $x_{jk}$ denote data at date $D_{jk}$. Then, the mean and variance are computed as

$$M = \sum_{k=1}^{N_Y}\sum_{j=-N_{1/2}}^{+N_{1/2}} w_j x_{jk} \tag{3}$$

$$V = \sum_{k=1}^{N_Y}\sum_{j=-N_{1/2}}^{+N_{1/2}} w_j (x'_{jk})^2 \tag{4}$$

where $x'_{jk} = (x_{jk} - M_j)$ denotes the anomaly about the climatological mean.

The probability distribution of anomalies about the mean is constructed with the aid of the family of distributions

$$p_\epsilon(x) = \sum_{k=1}^{N_Y}\sum_{j=-N_{1/2}}^{+N_{1/2}} w_j K_\epsilon(x - x'_{jk}) \tag{5}$$

where $\epsilon$ is considered to be a small positive number and

$$K_\epsilon(x) = \begin{cases} 0 & if\ |x| > \epsilon \\ (2\epsilon)^{-1} & if\ |x| \le \epsilon \end{cases} \tag{6}$$

The CDF is defined at the data values

$$\tilde{x} \in \{x'_{jk} \mid j = -N_{1/2},\ldots,N_{1/2}, k = 1,\ldots,N_Y\} \tag{7}$$

by integrating $p_\epsilon$ and then taking the limit $\epsilon \to 0$:

$$CDF(\tilde{x}) = \lim_{\epsilon \to 0} \int_{-\infty}^{\tilde{x}} p_\epsilon(x)\,dx \tag{8}$$

If the weights $w_j$ were all equal, the CDF would assume the values given by Hazen's plotting positions (cf. Wilks, 2006). For intermediate data points $x'$, the CDF is defined by linearly interpolating the probability between the closest data points enclosing $x'$. Therefore, quantile values are obtained through linear interpolation of the CDF at the data points. Note, that the definition of the CDF is consistent with the definition of the variance of anomalies in the sense that $\lim_{\epsilon\to 0} p_\epsilon$ will have the variance given by Eq. (4).

Statistics of the wave parameters (the significant wave height and the mean wave period) require extra care as due to the variable extent of the sea-ice there are no wave analysis data at sea points near polar regions in some years. Therefore only points where there are analysis values of at least 25% of the full time window available are taken into consideration; otherwise the grid-point value is set as missing.

# References


Jung T, Leutbecher M. 2008. Scale-dependent verification of ensemble forecasts. *Quart. J. Roy. Meteor. Soc.* **134**: 973-984.

Wilks DS. 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd edition. Academic Press: USA.