Additional information on the dataset groups and datasets in the ModE-RA Experiment

General information

The ModE project at WDC contains three components, each provided as a separate experiment (Figure 1). This additional information provides a description of the dataset group and datasets in the Mode-RA experiment. In another additional info file you will find R scripts for the quality control and assimilation of observations into the ModE-Sim ensemble using an Ensemble Kalman filter. For the underlying simulations (ModE-Sim) and the related reanalysis products (ModE-RAclim) equivalent info files can be found linked at their experiment page.

	ModE-RA	ModE-RAclim
Members	20	100
Prior	transient	time-invariant
Variability	>71-year: model response to forcing <71 year: assimilated observations	>71-year: NO variability <71 year: assimilated observations

Dataset groups in ModE-RA

The experiment ModE-RA contains one dataset group that refer to two subsets of the ModE-Sim ensemble, which serve as a prior for ModE-RA. In the earlier period ModE-RA is based on ModE-Sim Set 1420-3 and in the later period on ModE-Sim Set 1850-1. ModE-RA is not split into the two periods of the ModE-Sim prior, because the assimilated observational time series lead to a smooth transition between the two periods of the ModE-Sim sets.

Datasets within the dataset groups

The dataset group contains the following Datasets:

members (anomalies w. r. t. 1901 to 2000) - contains anomalies with respect to the period 1901-2000 for all 20 individual ensemble members of ModE-RA, called members 041 to 060. Files are provided as NetCDF files, with one variable from one ensemble member per file for the whole period. Filenames are structured as in the following example:

ModE-RA_m041_geopoth_50000_anom_wrt_1901-2000_1421-2008_mon.nc, where
- ModE-RA is the experiment name.
- m041 is the member number .
- geopoth_50000 is the variable name (geopoth) including, for 3D variables, the level (here: 500 hPa). A list of the variables can be found below.
- anom stands for anomalies.
- wrt_1901-2000 is the reference period for the anomalies.
- 1421-2008 is the period covered.
- mon means monthly time resolution.

ensemble statistics (anomalies w. r. t. 1901 to 2000) - contains ensemble means, ensemble standard deviations, and ensemble minima/maxima of all 20 individual ensemble members of ModE-RA. Files are provided as NetCDF files, with each file containing one statistical property computed for one variable and for the whole period. Filenames are structured as in the following example:

ModE-RA_ensstd_geopoth_50000_anom_wrt_1901-2000_1421-2008_mon.nc, where
- ModE-RA is the experiment name
- ensstd is the statistical property provided for the set (here: ensemble standard deviation).
- geopoth_50000 is the variable name (geopoth) including, for 3D variables, the level (here: 500 hPa). A list of the variables can be found below.
- anom stands for anomalies.
- wrt_1901-2000 is the reference period for the anomalies.
- 1421-2008 is the period covered.
- mon means monthly time resolution.

climatology of period 1901 to 2000 with model bias - contains the climatologies of the ModE-RA ensemble members and ensemble mean for all 12 months calculated over the the period 1901 to 2000. Files are provided as NetCDF files, with each file containing one variable. Note that adding back these climatologies to the anomalies leads to absolute values which contain the same model biases as ModE-Sim.

ModE-RA_ensmean_geopoth_50000_clim_wrt_1901-2000.nc, where
- ModE-RA is the experiment name.
- ensmean is the statistical property provided for the set (here: ensemble mean).
- geopoth_50000 is the variable name (geopoth) including, for 3D variables, the level (here: 500 hPa). A list of the variables can be found below.
- clim stands for climatology.
- wrt_1901-2000 is the period over which the climatology is calculated.

feedback archive - contains the metadata and information about preparation and assimilation of all observations considered in each assimilation step. It allows to track the impact of each observation on the reanalysis. It contains the following columns:

PRAOBSID	unique ID of each observation (e.g. monthly value)
PRARECID	unique ID of each series
NAME	name
LON, LAT	coordinates
ALT	altitude
YEAR_ASSIM	year in which it is assimilated
YEAR_OBS, MONTH, SEASON	year, month and season of the observation
VARIABLE	variable (e.g. sea level pressure)
UNIT	unit
TYPE	data type (e.g. instrumental measuement)
ORIGINAL_VALUE	original measurement
TRANSFORMED_VALUE	For instrumental data if the observation is in Julian calendar, it is interpolated to Gregorian. Otherwise, it is mostly a simple standardization in case of documentary and proxy data. Only in rare documentary data cases a log transformation or something similar is applied (see Reichen et al., 2022 for details)

CLIMATOLOGY_71YR	71-year moving climatology
ANOMALY_71YR	71-year moving anomaly
REF_PERIOD_START, REF_PERIOD_END	reference period
REF_DATA	the data set used for establishing the H operators
QC_INHERITED	quality flag inherited from the original source
OBS_ERR_VARIANCE	observation error variance
CYCLE	cycle in which the observation was assimilated
ACTIVE	TRUE if QC_INHERITED = 0, FALSE if it should not be assimilated but only added to the feedback archive for later comparison
T1, T2, T3, T4, T5, T6	coefficients of the forward operator H for 2m temperature in months 1 to 6 of each halfyearly file
RR1, RR2, RR3, RR4, RR5, RR6	coefficients of the forward operator H for precipitation in months 1 to 6 of each halfyearly file
Wetdays1, Wetdays2, Wetdays3, Wetdays4, Wetdays5, Wetdays6	coefficients of the forward operator H for wet days in a month in months 1 to 6 of each halfyearly file
SLP1, SLP2, SLP3, SLP4, SLP5, SLP6	coefficients of the forward operator H for sea level pressure in months 1 to 6 of each halfyearly file
U10_1, U10_2, U10_3, U10_4, U10_5, U10_6, V10_1, V10_2, V10_3, V10_4, V10_5, V10_6	coefficients of the forward operator H for 10m wind in months 1 to 6 of each halfyearly file
BGCHECK	For non-marine data the sqrt of the sum of the observation and model variances should be within 5 standard deviation. For marine data the limit is 2 standard deviations of the model.
ASSIM	whether the observation is assimilated (ACTIVE and BGCHECK both have to be TRUE)
LOC_LON1, LOC_LAT1, LOC_LON2, LOC_LAT2	length-scale parameters of the covariance localisation
BCKDEPMEM1, BCKDEPMEM2, BCKDEPMEM3, BCKDEPMEM4, BCKDEPMEM5, BCKDEPMEM6, BCKDEPMEM7, BCKDEPMEM8, BCKDEPMEM9, BCKDEPMEM10, BCKDEPMEM11, BCKDEPMEM12, BCKDEPMEM13, BCKDEPMEM14, BCKDEPMEM15, BCKDEPMEM16, BCKDEPMEM17, BCKDEPMEM18, BCKDEPMEM19, BCKDEPMEM20	background departure from each member (the first member MEM1 here refers to member m041 in the file names, … , MEM20 here to member m060 in the file names)
BCKDEPMEAN	mean of the background departures of all members
BCKDEPSPREAD	one standard deviation of the background departures of all members
ANADEPMEM1 ANADEPMEM2, ANADEPMEM3, ANADEPMEM4, ANADEPMEM5, ANADEPMEM6, ANADEPMEM7, ANADEPMEM8 ANADEPMEM9, ANADEPMEM10, ANADEPMEM11, ANADEPMEM12 ANADEPMEM13, ANADEPMEM14, ANADEPMEM15, ANADEPMEM16, ANADEPMEM17, ANADEPMEM18, ANADEPMEM19, ANADEPMEM20	analysis departure from each member (the first member MEM1 here refers to member m041 in the file names, …, MEM20 here to member m060 in the file names)
ANADEPMEAN	mean of the analysis departures of all members

BCKDEPMEM_SA1,, BCKDEPMEM_SA2,, BCKDEPMEM_SA3,, BCKDEPMEM_SA4, BCKDEPMEM_SA5,, BCKDEPMEM_SA6,, BCKDEPMEM_SA7,, BCKDEPMEM_SA8, BCKDEPMEM_SA9,, BCKDEPMEM_SA10, BCKDEPMEM_SA11, BCKDEPMEM_SA12, BCKDEPMEM_SA13, BCKDEPMEM_SA14, BCKDEPMEM_SA15, BCKDEPMEM_SA16, BCKDEPMEM_SA17, BCKDEPMEM_SA18, BCKDEPMEM_SA19, BCKDEPMEM_SA20	same as the background departure above but SA refers to the impact of an observation in the Serial Assimilation
BCKDEPMEAN_SA	mean of the background departures of all members in the Serial Assimilation
BCKDEPSPREAD_SA	one standard deviation of the background departures of all members in the Serial Assimilation
ANADEPMEM_SA1, ANADEPMEM_SA2, ANADEPMEM_SA3, ANADEPMEM_SA4, ANADEPMEM_SA5, ANADEPMEM_SA6, ANADEPMEM_SA7, ANADEPMEM_SA8, ANADEPMEM_SA9, ANADEPMEM_SA10, ANADEPMEM_SA11, ANADEPMEM_SA12, ANADEPMEM_SA13, ANADEPMEM_SA14, ANADEPMEM_SA15, ANADEPMEM_SA16, ANADEPMEM_SA17, ANADEPMEM_SA18, ANADEPMEM_SA19, ANADEPMEM_SA20	same as the analysis departure above but SA refers to the impact of an observation in the Serial Assimilation
ANADEPMEAN_SA	mean of the background departures of all members in the serial assimilation
NUMCELL	number of observations of same type per cell
NUMINCYCLE	number when the observation was assimilated in the given cycle
XSMOOTH	smoothed Hx (using the 71-yr climatologies as x)

The archive is organized in tabulator separated ASCII files for each half-yearly (Oct.-Mar., Apr.-Sep.) assimilation step.

List of variables in ModE-RA

The following variables are included in ModE-RA:

slp	mean sea level pressure [Pa]
temp2	2m air temperature [K]
totprec	total precipitation [kg m-2 s-1]
geopoth_50000	500 hPa geopotential height [m]
omega_50000	500 hPa vertical velocity [Pa s-1]
u_85000	850 hPa u-wind [m s-1]
v_85000	850 hPa v-wind [m s-1]

References

Reichen, L., Burgdorf, AM., Brönnimann, S., Franke, J., Hand, R., Valler, V., Samakonwa, E., Brugnara, Y., Rutishauser, T. (2022): A decade of cold Eurasian winters reconstructed for the early 19th century. Nat Commun 13, 2116. https://doi.org/10.1038/s41467-022-29677-8
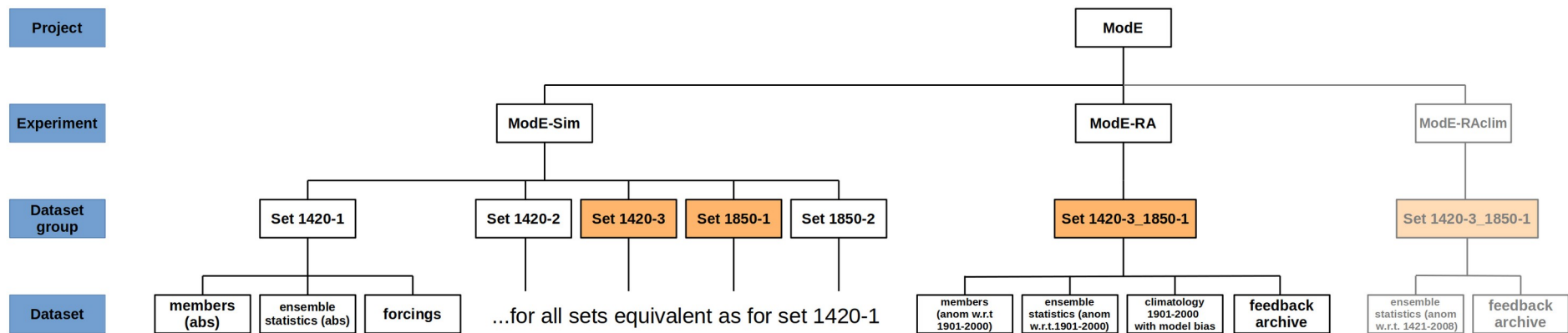
Figure 1: Overview of the structure of the ModE project at WDC. The ModE-Sim sets highlighted in orange are those which where used as a-priori state for the reanalysis product provided as separate experiments «ModE-RA»/»ModE-Raclim» within this project.