MASS2ANT Snowfall Dataset: Information for users

November 2020

N. Ghilain[1], S. Vannitsem[1], Q. Dalaiden[2], H. Goosse[2], L. De Cruz[1]
[1]Royal Meteorological Institute, Uccle, Belgium
[2]UCLouvain, Earth and Life Institute, Louvain-la-Neuve, Belgium

contacts RMI: stephane.vannitsem@meteo.be, nicolas.ghilain@meteo.be, lesley.decruz@meteo.be
contacts UCLouvain: hugues.goosse@uclouvain.be, quentin.dalaiden@uclouvain.be

We provide in this dataset maps at 5.5 km resolution of the daily and yearly accumulated snowfall over emerged land (Ice Sheet) of Dronning Maud Land (Antarctica) from 1850 to 2014. We used a statistical method to derive fine resolution maps from General Circulation Model (GCM) runs (CESM2, 10 runs). In the method, we searched for analogs in a database we constructed from the association between re-analyses large-scale meteorological fields (ERA5 and ERA-Interim) and RCM daily accumulated snowfall (RACMO2.3p5.5). RACMO2.3(p5.5) data are available freely on request (https://www.projects.science.uu.nl/iceclimate/models/antarctica.php). CESM2 CMIP6 runs are also freely available (https://esgf-node.llnl.gov/search/cmip6/). The complete description of the algorithm and performance is described in: *Ghilain N., Vannitsem S., Dalaiden Q., Goosse H., De Cruz L., Wei W., Reconstruction of daily snowfall accumulation at 5.5km resolution over Dronning Maud Land, Antarctica, from 1850 to 2014 using an analog-based downscaling technique*, submitted to Earth System Science Data (ESSD).

The MASS2ANT Snowfall dataset is composed of the annual estimations of snowfall over Dronning Maud Land, the daily time series for the total period for all the emerged grid points of the domain (2 files are given in example on zenodo, the all set is available on request to stephane.vannitsem@meteo.be), the principal component weights (PCs) time series and Empirical Orthogonal Functions (EOFs) offering the possibility to analyze the synoptic weather patterns associated to snowfall over the ice sheet and the PC time series from the re-analysis in case one wants to extend or improve the database. Realistic weather patterns can be recomposed in associating (product of matrices) the PCs with the EOFs. All files are stored in Netcdf4 format.

The method could be easily expanded to other sources (other GCM, other reference RCM or other reanalysis), and potentially for other parts of Antarctica. Samples of the programs used to generate this database are therefore also provided here.

A. The database files

1. Annual distributed snowfall accumulation at 5.5 km over Dronning Maud Land

File Names								
MASS2ANT_Snowfall_yearly_<Source1>_<Source2>_<Source3>_1850-2014.nc								
Number of files			Source 1	Source 2	Source 3	Topology	Grid	Resolution
10x2			CESM2r01, …, CESM2r10	ERA-I / ERA5	RACMO2.3	Maps	Stereo Polar	5.5 km
Field 1			Time frame		Name	Unit	Size	
Annual accumulated snowfall			1850 - 2014		‘Snowfall’	mm.w.e	430x200x165	
Field 2	Name	Unit	Size		Field 3	Name	Unit	Size
Latitude	‘Latitude’	degree	430x200		Longitude	‘Longitude’	degree	430x200

2. Daily accumulated snowfall time series over 1850 to 2014 for emerged points from RACMO2.3 grid at 5.5 km

File Names							
MASS2ANT_Snowfall_daily_<Source1>_<Source2>_<Source3>_1850-2014.nc							
Number of files			Source 1	Source 2	Source 3	Topology	Number of points	Resolution
2 (10x2 on request)			CESM2r01, …, CESM2r10	ERA-I / ERA5	RACMO2.3	Time series	46725	Daily
Field 1			Time frame		Name	Unit	Size	
Daily accumulated snowfall			1850 - 2014		‘Snowfall’	mm.w.e	46725x165x365	
Field 2	Name	Unit	Size		Field 3	Name	Unit	Size
Latitude	‘Latitude’	degree	46725x1		Longitude	‘Longitude’	degree	46725x1
Field 4	Name	Unit	Size		Field 5	Name	Unit	Size
Column RACMO	‘Column’	-	46725x1		Line RACMO	‘Line’	-	46725x1

3. Daily Bias Corrected Principal Components CESM2 time series over 1850 to 2014

File Names					
MASS2ANT_PCs_BC_<Source1>_<Source2>_1850-2014.nc					
Number of files	Source 1	Source 2		Topology	Resolution
2	CESM2	ERA-I / ERA5 (Z500, Rh700, SP, Precip)		Time series	Daily
Field 1	Time frame		Name	Unit	Size
Principal Components	1850 - 2014		‘PCs’	-	60225x40

4. Maps of 40 first Empirical Orthogonal Functions of main fields from reanalysis

File Names							
MASS2ANT_EOFs_<Source2>.nc							
Number of files			Source 2	Topology	Grid		Resolution
2			ERA-I or ERA5	Maps	Regular Lat - Lon		1 degree
Field 1				Name	Unit	Size	
Empirical Orthogonal Functions				‘EOFs’	-	57x51x40	
Field 2	Name	Unit	Size	Field 3	Name	Unit	Size
Latitude	‘Latitude’	degree	57x51	Longitude	‘Longitude’	degree	57x51

5. Daily Principal Components reanalysis time series over 1979 to 2000

File Names				
MASS2ANT_PCs_<Source2>_1979-2000.nc				
Number of files	Source 2		Topology	Resolution
2	ERA-I or ERA5 (Z500, Rh700, SP, Precip)		Time series	Daily
Field 1	Time frame	Name	Unit	Size
Principal Components	1979 - 2000	‘PCs’	-	40x8036

B. The programs

Program performing the construction of analogs database and the estimation on model time series
Fortran90, use of OpenMP Libraries for parallel computing

Inputs: Snowfall_2D (2D): the RCM daily snowfall accumulation [nb_days_obs x nb_pixels]
PCs (2D): PCs (GCM members & reanalysis) [nb_PCs x (nb_days_model x nb_members + nb_days_obs)]

Output: - Ypred_Precip_New → one ascii file per pixel with snowfall time series

```
use OMP_LIB

do pixel=1,nb_pixels
        !Downscaling by the method of the analogs:
        !-----------------------------------------------
        AnalogueNumber=20
        Snowfall=Snowfall_2D(:,pixel)
        allocate(Ypred(nb_days_model*nb_members+nb_days_obs),v_ascend(AnalogueNumber))
        !$OMP PARALLEL DO default(private) shared(Ypred, PCs, AnalogueNumber, Snowfall, IndTrain)
        do j=1,nb_days_model*nb_members+nb_days_obs
                r=norm2(PCs(:,IndTrain)-transpose(spread(PCs(:,j),dim=1,ncopies=nb_days_obs)),dim=1)
                mk = .true.
                do i=1,AnalogueNumber
                        v_ascend(i)=minval(minloc(r,mask = mk))
                        mk(minloc(r,mask = mk))=.FALSE.
                enddo
                Ypred(j)=sum(Snowfall(v_ascend))/AnalogueNumber
        enddo
        !$OMP END PARALLEL DO
        deallocate(v_ascend)

        !Scaling
        !Correction on Reanalysis: 1. linear correction, 2. Quantile mapping
        !-------------------------------------------------------------------------
        Train_ERA50=Ypred(1+nb_members*nb_days_model:size(Ypred))
        Train_Obs=Snowfall
        deallocate(Snowfall)

        call QM_correction(Train_ERA_50,Train_Obs,Train_ERA5_c)

        !Correction on each CESM2 member compared to ERA5
        !---------------------------------------------------------------
        allocate(Forecast10(nb_days_model,nb_members))
        do j=1,nb_members
                Forecast0(1:size(Forecast0))=Ypred(1+(j-1)*nb_days_model:j*nb_days_model)
                call QM_correction(Forecast0,Train_ERA5_c,Forecast_c)
                Forecast10(:,j)=Forecast_c
        enddo

        !Store the result and print the time series
        !-------------------------------------------
        allocate(Ypred_Precip_New(1,nb_days_model*nb_members))
        Ypred_Precip_New=reshape(Forecast10,(/1,nb_days_model*nb_members/))
        file_out="SF"//num2str(iter)//"_1850_2014_CESM2_ERAI_5km_Member1to10.txt"
        open(fileID,file=file_out)
        write(fileID,'(F12.5)') (Ypred_Precip_New(1,hh),hh=1,size(Ypred_Precip_New,2))
        close(fileID)
        deallocate(Ypred_Precip_New)
```

```
enddo
```

Subroutines used by the main program

subroutine sort(A,B,D,NB)

```
real, dimension(:), intent(in) :: A
integer, intent(in) :: D, NB
real, dimension(NB), intent(out) :: B
logical, dimension(:), allocatable :: mk

allocate(mk(size(A)))
mk = .TRUE.
if (D==1) then
        do i=1,NB
                B(i)=minval(A,mk)
                mk(minloc(A,mk))=.FALSE.
        enddo
else
        do i=1,NB
                B(i)=maxval(A,mk)
                mk(maxloc(A,mk))=.FALSE.
        enddo
endif
deallocate(mk)
```

end subroutine sort

subroutine QM_correction(A,B,A_corr)

```
threshold = 0.18; denom=10.0; total=80.0
call sort(A,A_s,2,denom)
maxA=sum(A_s)/denom
call sort(A,A_s,1,denom)
minA=sum(A_s)/denom
call sort(B,B_s,2,denom)
maxB=sum(B_s)/denom
call sort(B,B_s,1,denom)
minB=sum(B_s)/denom
A_temp=(A-minA)/(maxA-minA)*(maxB-minB)+minB
where (A_temp<threshold) A_temp=0.0
nobnd = floor(total*denom)
allocate(bound(nobnd+1),c1(nobnd),c2(nobnd),cs1(nobnd),cs2(nobnd))
bound(1)=0
do i=2,nobnd+1
        bound(i)=-(1.0-i)/denom
        c1(i-1)=count(A_temp>=bound(i-1).and.A_temp<bound(i))
        c2(i-1)=count(B>=bound(i-1).and.B<bound(i))
        cs1(i-1)=sum(c1(1:i-1))
        cs2(i-1)=sum(c2(1:i-1))
enddo
cs1=cs1/sum(c1)
cs2=cs2/sum(c2)
A_corr=A_temp
do r0=1,nobnd
        correc=minval(minloc(abs(cs2-cs1(r0))))
        call random_number(randmat)
        where (A_temp>=(r0-1.0)/denom.and.A_temp<r0/denom)
                A_corr=max(A_temp-(r0-correc+0.5-randmat)/denom,0.0)
        end where
enddo
deallocate(c1,c2,cs1,cs2)
```

end subroutine QM_correction

Program performing the EOF decomposition and Corrections to PCs

Matlab, use of Climate Toolbox function eof (Chad Greene)

```
%Grid definition Reanalysis
lon=-10.0:1.0:60.0; lat=-90.0:1.0:-40.0; [lonE,latE]=meshgrid(lon,lat);

%Grid definition GCM
lonC=-10.0:1.25:60.0; latC=-90.0:1.0:-40.0; [lonC1,latC1]=meshgrid(lonC,latC);

VarName={'Geopotential height at 500hPa','Relative humidity at 700hPa','Surface Pressure','Log(Daily Precip
+1)'};

for Var=1:4
   Variable = Fields(:,:,:,Var);
   %Decomposition into Empirical Orthogonal Functions (EOFs)
   [eof_maps,pc,expv]=eof(Variable,nb_modes); % Geopotential height at 500hPa
   %Regridding of nb_modes (=10) first modes of the 4 fields from Reanalysis to GCM grid
   for i=1:nb_modes
     erv=eof_maps(:,:,i); EOFr0=griddata(lonE,latE,erv',lonC1,latC1); EOF_ERAr(:,:,i,Var)=EOFr0';
   end
   PC_ERA(:,1+(Var-1)*nb_modes:nb_modes*Var)=pc;
end
ERA=reshape(EOF_ERAr,nb_col*nb_lin,nb_modes,Var);

%Linear correction of GCM
for member=1:nb_member
  for Var=1:4
    [eof_short,pc_short,expv]=eof(Var_short,nb_modes);
    [eof_long,pc_long,expv]=eof(Var_long,nb_modes);
    CESM=reshape(eof_short,nb_col*nb_lin,nb_modes);
    for bb=1:nb_modes
      PO=polyfitn(CESM,ERA11(:,bb),1);
      PCC10_corr(:,bb)=(PO.Coefficients(nb_modes+1)+PO.Coefficients(1:nb_modes)*pc_long)';
      clear PO;
      %Extra correction: CDF matching
      Corr_PC(:,bb+(Var-1)*nb_modes)=CDF_matching(PCC10_corr(:,nb_modes),PC_ERA(:,nb_modes));
    end
  end
end

PCs_to_store=[Corr_PC; PC_ERA];
```