Stations:

All stations, except for Terranova Bay are part of the BSRN (Baseline Surface Radiation Network) and collect radiation measurements following specific protocols to yield high quality data. Terranova Bay is a candidate station for the network and therefore has started collecting measurements but I would say they still need to fix a few things. Generally Concordia and South Pole often show clear sky conditions, whereas the rest of them show all three sky conditions (clear, overcast and mixed). Station clusters could be Concordia and South Pole vs coastal stations (the rest).

"acronimo"	Stazione	
dom	Concordia	2015-2016
jbogo	Terranova Bay	2018-2022
neu	Neumayer	2015-2016
sho	Showa	2015-2016
spo	South Pole	2015-2016

Header description

		Value range
Date		
SWD	Downward shortave radiation (or global shortwave radiation))	0 / ~1500 W/m^2
DIF	Diffuse shortwave radiation	0 / ~500 W/m^2
SZA	Solar Zenith Angle	0 – 90° = sun above horizon > 90° = sun below horizon
SWlabel	Label for sky conditions, derived from RADFLUX algorithm based on SW data	CS = clear sky conditions Cl = cloudy sky conditions OC = overcast sky conditions
LWD	Downward longwave radiation	50 / 400 W/m^2
LWU	Upward longwave radiation	100 / 500 W/m^2
Ts	Surface air Temperature	-90 / 10°C
RH	Relative Humidity	0 – 100
Ratio	Water Vapor Pressure/Ts	0 - 5
LWlabel.rf	Label for sky conditions, derived from RADFLUX algorithm, based on LW data	CS = clear sky conditions Cl = cloudy sky conditions OC = overcast sky conditions
LWlabel.VdB	Label for sky conditions, derived from Van den Broeke algorithm, based on LW data	CS = clear sky conditions Cl = cloudy sky conditions OC = overcast sky conditions

Important notes

- Ratio is derived from a parametrization that involves both Ts and RH;
- Labels derived from shortwave radiation (SW, solar) can only be computed during daylight (= austral summer), but are generally more reliable because we can tell from the behaviour of the data if the label is correct;
- Labels derived from longwave radiation (LW, terrestrial/thermal) can be computed year-round (= also during polar night) but from the data we cannot tell if the label is correct, unless in specific cases;
- Labels are obtained by setting thresholds for the value of cloud fraction (output of my models) for SW based methods, or value of a clear sky index (output of LW based methods). Therefore we can cluster data differently according to the thresholds;
- Know bias: LW based algorithms are blind to cold, high clouds (like cirrus clouds), therefore if they are the only type of clouds SW based methods will correctly identify the sky as cloudy, but LW based methods will identify them as clear;
- NaN values for missing data or physically impossible values that were deleted from the data frame
- SWD, LWD and LWU, Ts and RH are independent variables, the rest are dependent:
  DIF is part of SWD, ratio is obtained from Ts and RH, SZA depends on date time and longitude and latitude