<u>**README - Interpolation**</u>

Author: James McCreight - mccreigh@nsidc.org

The purpose of this document is to describe the way the original ERA40 data was treated when regridding in space and time.

Ten variables were chosen for reinterpolation from the original ERA40 Northern Hemisphere, 6 hourly data set. The variables chosen are typical drivers needed for land(& sea) surface models. The list of the variables:
(instantaneous:)
2 meter air temp
2 meter dew point temperature
10 m U direction wind component
10 m V direction wind component
Surface Pressure
(accumulated:)
Surface Downward Shortwave Radiation
Surface Downward Longwave Radiation
Large Scale Precipitation
Convective Precipitation
Snowfall

The first 5 variables are considered instantaneous values at the given time while **the flux variables are given by ECMWF as 6 hour accumulations beginning at the observation time**.

First, the variables are interpolated in space to the EASE100 northern hemisphere (Nl.gpd) grid using at least 2 different interpolation schemes. After the data is regridded in space, this new data is then interpolated in time.

<u>Interpolation in Space</u>

Two interpolation schemes were selected to regrid the data in space. Bilinear interpolation according to ECMWF specifications was followed as closely as could be reconstructed from the information available at their website. Cressman interpolation with a variable radius of interpolation was also chosen. Nearest Neighbor interpolation is also available as a special case of bilinear interpolation.

Bilinear Interpolation:

The weights and bounding neighbors were computed beforehand (offline) in the idl routine make_bilinr_box.pro.

In bilinear interpolation there are really 3 interpolations going on, 2 in lon and 1 in lat. First, 2 points of equal latitude on the N80 grid are being interped to a temporary pt between them which has the same longitude as the EASE100 point that will eventually be interpolated to. This occurs twice, once for the bounding pair north of some EASE100 pt and once for the bounding pair south of that point. Then a secondary interpolation occurs in latitude from these two temporary points to the actual EASE point.

ECMWF metions a weighting that is based on a land surface mask. This weighting applies only to the air and dew point temperatures and to the convective precip, the variables that depend on surface type. So, a land-sea mask must be considered. We define a point as land if there is greater than 50% land in the cell represented by the grid point. If a cell contains 50% or less land, it is considered a water point. In each individual interp (just in lon or just in lat), 0,1, or 2 N80 points may differ from the ease point in surface classification type, either land or water. If the class is the same for both N80 and the ease pt then nothing happens. If ONLY one differs in classification, the weight on that point gets reweighted by .2 and the weight on the other point in the interp gets rewighted so that the equation remains 'balanced', ie (.2)(orig weight pt a)+(?)(orig weight pt b)=1. In the last case, if both N80 points in a lon interp have a different class than the ease pt then no weights are applied in the lon interp but a weight of .2 is applied in the lat interp that follows the lon interp. The only case left to discuss: If all 4 N80 points disagree in lsm class with the ease pt, ie if both the temp lon interps are supposed to be weighted in the lat interp. In this case nothing is done, there is no reweighting.

Inter-twining this reweighting with the bilinear interp eqns isnt too difficult but does take some time and can cause some confusion when reading the code. Here is the full explaination:
We are interpolating in longitude between the points A and B (on either the upper or lower edge of the bounding box) to the temporary interpolation point x.The original equation is
$x = A + w \cdot (B - A) = A(1 - w) + Bw$
where w is the bilinear weight computed offline. Now, there are two possibilities when one of the points (but not both) has a different land-sea value from the point we are interpolation to (this is not x). Suppose B dosent match, then we have the equation
$x_B = .2 \cdot A(1 - w) + \ ? \cdot B \cdot w.$
We can solve for ? via the equation
$1 = \ ?(1 - w) + .2w$
with the result
$? = \frac{1\text{-}.2w}{1 - w}.$
Substituting this into the equation for x_a we get
$x_B = A + .2 \cdot w \cdot (B - A).$
If we suppose the other case, where B dosent match in surface type, then we follow similar steps and get
$x_A = .2A + .8B + .2 \cdot w \cdot (B - A).$
Now, since I felt tricky, there is a way of combining these equations into a single equation using appropriate "mask weights", msk. The equation is
$x = msk1 \cdot A + (1 - msk1) \cdot B + msk2 \cdot w \cdot (B - A)$
where
If a disagrees in surface type then msk1=msk2=.2
If b disagrees in surface type then msk1=1 & msk2=.2
Otherwise, both "mask weights" are set to 1 and we simply have the original equation.
Recall that if both points agree or disagree in surface type with the new point being interpolated to then there is no weighting until the interpolation in latitude. This same eqation is used for that interpolation and the "msk weights" are set in the same way for the latitude interpolation.

Interpolating precip, which often involves small amounts and zero amounts in original neighboring cells, can create a drizzle affect where many new cells will have a very slight amount of rain even though they should probably be zero. The ECMWF addresses this problem and prescribes a condition and threshold at which precip amounts should be set to zero. If the interpolated value or the nearestneighbor of the interploated point are less than .00005, set the interpolated value to 0.

There are several places on the EASE grid where no N80 bounding box existed on the northern hemisphere EASE grid (near poles and equator), in these cases nearest neighbor interpolation is used. As mentioned above, nearest neighbor interpolation may be done as a special case of bilinear interpolation. To do this one must create a bilinear weights file in which each of the corner points of the bounding box is set to the nearest neighbor. This file is not available at the current time.

Cressman Interpolation: This is the standard cressman scheme. The distances to each of the nearest 50 N80 neighbors of each EASE point have been calculated offline. In the code, the user may choose the cressman radius and the N80 points in side this are used. If there are no points inside the radius the error value of 9999 is used.

The same precipitation threshold as was used for the bilinear scheme is also applied for the cressman scheme.

Temporal Interpolation

The goal of interpolating the variables in time, down to 3 hourly resolution, was to make the data either "instantaneous" at the time stamp or to make it a 3 hour average centered at the time stamp. For the instantaneous variables this is easy, simple linear interpolation in time was used to find values midway between 6 hourly values.

It was desirable to output specific humidity at the 3 hourly resolution. This was done using the tetens equation after interpolating the needed variables (pressure, temperature, & dew point temperature) to the 3 hourly resolution. The tetens equations for vapor pressure and saturation vapor pressure:

$$e = (estr) \cdot exp\left(b \cdot \frac{T_d - T_{tr}}{Td - T_{tet}}\right) \qquad es = (estr) \cdot exp\left(b \cdot \frac{T - T_{tr}}{T - T_{tet}}\right)$$

the RH and Sh eqns

$$RH = \frac{e}{es} \qquad SH = \frac{e \cdot \frac{\varepsilon}{p - e}}{1 + e \cdot \frac{\varepsilon}{p - e}}$$

where
T_d is 2m dewpoint temp
T is 2m temp
p is pressure
$b = 17.2694$ - 1/K
$estr = 611.73$ - Pa these two are the tripple point p and T
$T_{tr} = 273.16$ - K tripple pt temp
$T_{tet} = 35.86$ - K
$\varepsilon = .622$ - Psycometric Constant
You may ask why the realtive humidity is included here, that will be answered shortly.

Temporal interpolation is done with an eye towards conserving the accumulations but expressing them as averages. For the fluxes, which are accumulated values over the 6 hour period beginning at the time of observation, the situation is more complex. First, the accumulated values must be partioned in a meaningful way and so that the accumulations can be centered about 3 hourly steps, which means dividing each 6 hour period into 4, 1.5 hour, pieces. Then these pieces must be combined and averaged over 3 hours to give average values centered at the 3 hourly timestamp.

In particular, partitioning the solar radiation requires some detail because the amount of solar radiation varies nonlinearly in time. This description will be left till later. For the other flux variables, the interpolation is more straight forward and each of these variables is treated in the same way. While the variables might each vary nonlinearly in time, we have no real way of reconstructing the variations in a very accurate way. For each of these, we could simply choose to break up the accumulations equally over the 6 hour periods. However, if we consider each variable, we can recognize that there is some correlation of each to the amount of moisture in the atmosphere. The precipitation variables are directly related to the amount of moisture in the atmosphere as is the longwave radiation, in some crude fashion. Therefore, to reconstruct some variation in the flux variables in time, we choose to partition the 6 hour accumulations of these variables into 2 parts, each weighted by the amount of relative humidity at the beginning of each 3 hour period inside the 6 hour accumulation. This is why relative humidity was caluclated along with specific humidity above. Relative humidity is better suited for this calculation because it varies more with temperature and should help bring the diurnal cycle into the redistribution of the accumulations.???other reasons??? The size of the 2 parts that the accumulation is broken into is determined by the relative humidity at the beginning and middle of the 6 hour period (as the original data is cumulative for 6 hours beginning at the timestamp). The ratio of each relative humidity value to their sum determines the weight on the 6 hour accumulation given to each of the two pieces. As the data is to be recentered about the timestamps, we have to break the 6 hourly accumulations into 4 pieces. To do this, we simply half each of the two pieces we currently have. Now, two pieces lie to each side of every timestamp, these pieces are added together and their sum is divided by 3*3600s to find and instantaneous (1/s) average for the 3 hour period centered at the time stamp.

Since the solar radiation varies nonlinearly over the 6 hour accumulation, breaking it up in the same manner as the other flux variables will not work and, in this case, we have other information that will help us reconstruct how the solar varied in time over each 6 hour period. What we know is the cosine of the solar zenith angle at any point on the earth for a given day of the year. So, offline, a set of weights is calculated for partitioning accumulated solar radiation by integrating the cosine of the solar zenith angle (at every EASE100 grid cell location) over 90 minute periods. Just as for the other variables, if we are to center the interpolated data about the timestamp, we must break the 6 hour periods into 4, 90 minute pieces. Since we have the integrals every 90 minutes for the cos of the solar zenith angle, we find the ratio of each 90 min piece to the sum of all the 90 min pieces in the 6 hour period in which it resides to find the percentage of the 6 hourly accumulation that occured in that 90 minutes. Summing both 90 minute periods on each side of every time stamp and dividing by 3*3600s we find the average instantaneous flux for that 3 hour period.

The solar radiation is also weighted by relative humidity but in a slightly different way than the other variables. First, we expect solar at the surface to vary inversely with amount of moisture in the atmosphere. This means that the weights "switch places" as compared to the other variables. Also, because the accumulations were already broken up according to the integrals of the cos of the solar zenith angle, we dont have to split up the 6 hourly accumulations with the relative humidity. The result is that we calculate the weight of the relative humidity for the first 3 hour period in each 6 hour period as above. Then we subtract this weight from 1 and take its ratio with .5 to get a reweighting on the solar for the timestamp corresponding with the beginning of the 6 hour accumulation period. To conserve the total amount of solar we then must take the amount that we added/subtracted from this value by reweighting it and subtract/add it to the other solar value within the 6 hour accumulation.