Ensemble Pre-Processor Program (ENS_PRE) User Manual

1. Description of the Algorithm

In order to integrate the Long Range Climate Prediction Center (CPC) forecasts of precipitation and temperature, the probability anomaly shift described by the CPC forecasts must be converted into a precipitation or temperature shift. That is, the probability anomalies must be translated into real physical temperature or precipitation measured in degrees or millimeters.

This is accomplished by computing the shape of the new distribution as it is described in the CPC forecasts and then mapping the values in the MAP or MAT time series from the old distribution to the new distribution. The ideal mapping would determine the climatological probability of a particular value, then extract the value associated with that probability from the new forecast distribution. In this way, the new forecast distribution would be reflected in the MAP or MAT time series.

However, the computational requirements of such a scheme would be large and the simpler method of computing the shift in the medians of the distributions and then using this shift for all values of the time series is used to simplify the calculations. In the tails of the distributions the shift is either too large or too small depending on the shape of the climatological distribution's tail and the shape of the forecast tail, but this discrepancy is insignificant in light of the forecast accuracy and information content. For the temperature time series an additive adjustment is computed and for the precipitation time series a multiplicative adjustment is computed. A different adjustment is computed for each day of the year.

The algorithm produces invalid values of precipitation and temperature when the forecast shift is greater than 30 percent. Therefore, the program checks the value of the forecast shift, and if the forecast shift is greater than 30 percent it resets the forecast value to 30%.

The adjusted time series are not permitted to wrap around a year. The reason for this limit is a result of the output CARD file format. It is not possible to store the same month twice in the CARD format consequently it is not possible to adjust the same month twice. The inability to use 13 month adjusted time series is unlikely to create undue hardship for hydrologic forecasts as the skill in such long forecast is likely to be slight. As new database formats become available, 13 month adjusted time series can be made available as well.

The CPC forecasts are generated over an area that is much larger than the usual RFC basin. The CPC forecasts are on a scale that is on the order of a 4 degree by 4 degree grid. In order to have the climatology that is input to the **ens_pre** program and the CPC forecasts coincide, the climatology that is used as the parametric input to the program should be aggregated up to some suitable collection of basins.

The ens_pre program will read a grid of the CPC forecasts. For each area a centroid is provided and the value at the centroid is used to represent the entire area. This assumption is validated by the smooth variation in the forecast values.

2. Program Setup

2.1 Calibration

The program **ens_pre** requires input parametric information in the form of climatological distributions for a suitable area. The basin averaged MAP and MAT time series can be aggregated to a larger basin using the MCP program and the WEIGH-TS operation to create an area weighted average of the sub-basin MAP and MAT time series. Once the aggregate time series has been generated, the climatological distributions can be extracted.

The program **ens_pre_cp** (ensemble preadjust calibration) is used to calibrate the preadjust process by computing statistics used to make the adjustments.

2.2 Directory Structure and Apps_defaults Tokens

The program uses the following directories:

For control input files:
$(ens_dir)/input/$(ofs_level)/ens_pre

For parametric files including the tabular file of CPC forecasts:
Token: preadj_dir
$(ens_dir)/files/$(ofs_level)/cpc_fcsts
and
$(ens_dir)/files/$(ofs_level)/cpc_fcsts/stats

For NCEP grid files:
Token: enspre_griddb
$(FXA_DATA)/Grid/SBN/netCDF/CONUS211/CPCoutlook

For the output precipitation and temperature files
Token: preadj_outts_dir
$(calb_area_ts_dir)/pre

The OFS FCST technique PREADJ causes the fcst program to add the "/pre" to the calb_area_ts_dir token value, facilitating the use of the adjusted time series.

2.3 File types and Formats

The control input file is the only file that should be edited. The parametric files, the CPC forecast files, the record of the adjustment factors, all of the files that are located in the parametric files directory are generated automatically and are fixed format. DO NOT EDIT THE PARAMETRIC FILES.

2.4 Control Input Options

The control input file has the following format and options.

The general format is the same one used in ESPADP batch files: token=value. Token is the option and the value is the value assigned to that option. There is no blank space permitted in the string.

All of the input options are required. The input options are

1. forecast_file - The name of the file that will hold the CPC forecasts.
2. runtype - segment, fgroup or cgroup
3. run_area_id - The name of the area being run which is used as a prefix for the parameter files.
4. dummy_dytpe - The data type that is dummied out in the ESP time series definitions (either MAP or MAPX). If program **ens_pre** knows the dummy data type name, then it will skip those data types in its processing making the whole process a little faster.
5. centroid_lat - The latitude of the centroid of the area.
6. centroid_lon - The longitude of the centroid of the area.

Comments are allowed. Comments are lines preceded with a #.

The following is a sample input deck:

```
# an input deck for the ens_pre program
forecast_file=DES.cpc
runtype=fgroup
run_area_id=DES
dummy_dtype=MAPX
centroid_lat=45.00
centroid_lon=-124.91
```

2.5 Parametric File Descriptions

Within the parametric files there are 4 file types: climate statistics files, CPC forecasts, a record of the adjustments and a log that tracks erroneously high adjustments. There may be multiple versions of each set of 4, for example, one for each different forecast area.

In the directory $(ens_dir)/files/$(ofs_level)/cpc_fcsts there will be three files and one directory.

The three files will be
the CPC forecasts - for example: DES.cpc
the record of the adjustments – for example: DES.old
the tracking file - for example: DES.correct

The one directory stats will hold the climate statistics generated in the calibration process.
$(ens_dir)/files/$(ofs_level)/ens_pre/stats/
for example: DES_precip.stat DES_temp.stat

Note in these example files that each file begins with the same three letter prefix. This is the run_area_id that is input into the control input file.

The CPC forecasts are entered via the GUI in ESPADP using the units that are specified in the GUI. The .stat files are the product of the calibration process done with program **ens_pre_cp** and the adjustment files are generated by program **ens_pre**.

2.6 Control Output File

The diagnostic/control output file name and location are user selected and are written to the users ens_output directory. The values read from the grids and used in the time series adjustment for the CPC anomaly forecasts are printed in to the output file.

3. Program Operation

3.1 Program Execution

The script **ens** can be used to run ens_pre:

ens -p ens_pre -i <input_file> -o <output_file>

The program parses the input file and then reads from the ESPPARM file the names of the time series to be adjusted. These time series are then read by program **ens_pre**, adjusted according to the appropriate adjustment and then written out to a new directory structure that is parallel to existing calibration directory structure. The three letter directory pre is inserted after the $(calb_area_ts_dir) in the pathname/filename. All required directories are created by program **ens_pre**.

When running **ESP**, one may use the **PREADJ** technique to direct **ESP** to look in the directory $(calb_area_ts_dir)/pre for the card time series used as input.

All the time series that are used in the basins selected in the control input file are replicated by program **ens_pre**. Those that are not adjusted are linked to the originals. In this way, ESP can run against the whole new set of files without having to look sometimes in one location and the next time in another location.

3.2 Use of the gridded CPC files

The ens_pre program will read grids of the CPC anomaly forecasts. The grid files are stored in the directory indicated above and then read by program **ens_pre** at execution time. The centroid of the area being run is used to find a representative value of the forecast for the entire area. The centroid is input via the control file. If no centroid is provided, the program will extract the CPC forecast information from the forecast file generated by program **espadp** and referred to above as the CPC forecast file.

The program reads the 1-5 and 6-10 day forecasts from the"CPC forecast" file. If those forecasts are valid for the current run they will be used. If they are not valid they will not be used.

The start date of the preadjust run is also read from the"CPC forecast" file. It may be updated with program **espadp**.

The program **ens_pre** compares the dates of the forecasts in the grid files and the dated of the CPC forecasts in the"CPC forecast" file and uses the more recent of the two. If the grids and the"CPC forecast" file have the same date the"CPC forecast" file is used. In this way the user may override the grids simply by entering alternate values into the"CPC forecast" file. To make sure the grids are being used, the CPC forecasts in the “CPC forecast” file should be set to begin several months in the past.