Running ENS_POST

The Ensemble Post-Processor adjusts ESP output traces in order to account for biases and errors in the hydrologic models and their calibration and to account for uncertainty inherent in the models.

The post-processor accomplishes its goal through the application of an error model (described in Appendix A) to each individual trace.

1. Configuration

ENS_POST uses apps-default tokens to determine the location of directories required to operate.

1.1 Environment Variables

The following apps-defaults tokens must be specified for ENS_POST to properly execute:

Name	Example Implementation Directory	Purpose
ens_input	$(ens_dir)/input/$(ofs_level)	directory containing all input directories for the ens script
ens_output	$(ens_dir)/output	directory of output files generated by ens script
espts_dir	$(ens_files)/espts	directory of ESP output time series

2. Batch Mode

ENS_POST is intended to run from a batch command file via the ens script.

2.1 Running Batch Operations

The following can be used to run ENS_POST from a batch file:

```
ens -p ens_post -I <batch_filename> -o <output_filename>
```

where   batch_filename is the name of the input file to use with ENS_POST (described below)
            output filename is the file to contain the diagnostic output from the ENS_POST run

The input file must be located in the directory $(ens_input)/error_model and the output file is placed in $(ens_output)/$(LOGNAME).

2.2 Instructions

This section is intended to provide step-by-step instructions for a first time user to construct an ENS_POST batch file. All commands and actions, denoted in **bold**, will be described in sections 2.5 and

2.6.

Step 1: Set the Directory Containing ESP Output Time Series Files
Typically, this directory is that corresponding to the apps-default token espts_dir. However, if the user desires, they may manually set this directory by calling the command **ESPTS_DIR**.

Step 2: Set the Observed Segment Identifier and Data Type
The ESP output time series file must be for a segment within the OFS data base. There will normally be observed data stored in the OFS data base for the segment and for the same data type as the ESP time series output file. ENS_POST will try to read the observed data at run time if either ER2D or ER2S is the error model used. Specify the segment id and data type to use by calling the commands **OBS_SEGMENT_ID** and **OBS_DATA_TYPE**.

Step 3: Set the Type of Error Model to Use
ENS_POST will apply an error model to the ESP output time series. The user can specify which error model to apply from these three: ER1, ER2D and ER2S. Call the **ERROR_MODEL** command to set the error model.

Step 4: Run the Ensemble Post-Processor
To execute the post-processor, call the action **RUN_ENS_POST** giving it a value equal to the name of the ESP output time series file you wish to apply it to.

2.3 Batch File Format

The format of the batch file is the same as that of the ESPADP batch file, as described in section 3.3.3 of the ESPADP User Manual.

2.4 Definitions

This section provides definitions used throughout this section and the remaining sections. Additional definitions to those in this section will be provided as needed. The definitions are as follows:

<u>Command</u>: A batch input that sets a parameter for an ENS_POST action. The name of the command is identical to the <token> in the line of the batch input file.

<u>Action</u>: A batch input that instructs ENS_POST to produce output of some kind. There is only one available action currently within ENS_POST and that is **RUN_ENS_POST**.

2.5 Batch Commands

This section provides an alphabetical listing of all of the available batch commands. Commands, denoted in **bold**, set parameter values for an ENS_POST batch action (see Section 2.6). They do NOT result in output. The <value> within the line of the batch input file that contains the command is usually restricted, depending on that command. Acceptable values will be listed for each command. The following are ENS_POST commands:

ERROR_MODEL = <value>
Description: Sets the error model applied within ENS_POST. The available error models are described in Appendix A.

Acceptable Values: "ER1", "ER2D", "ER2S"
Default Value: N/A. If not specified, ENS_POST will fail to run.

ESPTS_DIR = <directory>
Description: Sets the directory ENS_POST searches for the ESP output time series file specified in the
    RUN_ENS_POST action.
Acceptable Values: A complete directory name.
Default Value: The directory pointed to by apps-defaults token espts_dir.

OBS_DATA_TYPE = <value>
Description: Sets the data type of the observed time series stored in the OFS database that will be
    searched for the latest observed value. ER2D and ER2S depend on this value and, when it is not
    found, will assume ER1 is applied for the first time step of the ESP output time series.
Acceptable Values: A valid OFS data type, typically "QINE" or "SQIN"
Default Value: N/A. If not specified, ER1 is applied for the first time step.

OBS_SEGMENT_ID = <value>
Description: Sets the segment id of the observed time series stored in the OFS database that will be
    searched for the latest observed value. ER2D and ER2S depend on this value and, when it is not
    found, will assume ER1 is applied for the first time step of the ESP output time series.
Acceptable Values: A valid OFS segment id of no more than eight characters.
Default Value: N/A. If not specified, ER1 is applied for the first time step.

2.6 Batch Actions

Actions, denoted in **bold**, instruct ENS_POST to produce some output using the current batch mode
command settings. For actions, the value on the batch input line defining the action is a list of options for
that action. The list of options is comma separated and must NEVER include spaces. The parameters are
given below for each action. The following is the ENS_POST action:

RUN_ENS_POST = <filename>
Description: Apply the specified error model to the ESP output time series file specified by the filename.
Options:
    1.filename: Must be the name of an ESP output time series file within the directory given by the
    current **ESPTS_DIR** command setting.

    *NOTE: Certain files must be available for ENS_POST to run: the parameter file (Appendix B) and
    the CDF files (Appendix C). These must be located within the $(ens_input)/error_model directory.
    The parameter file must have a name consisting of the first four components of the ESP output time
    series filename with a ".param" extension. This file must then contain within it the naming structure
    of the CDF files. See Appendix B for more details.*

2.7 Example

Figure 1.1 provides an example of an ENS_POST batch file.

```
# 1. Set the value of the ESPTS_DIR command setting.
ESPTS_DIR = /awips/hydroapps/rfc/nwsrfs/ens/files/oper/espts

# 2. Set the ERROR_MODEL.
ERROR_MODEL = ER2S

# 3. Set OBS_SEGMENT_ID and OBS_DATA_TYPE for the first ENS_POST run.
OBS_SEGMENT_ID = EMYSW
OBS_DATA_TYPE = QINE

# 4. Perform the first run.
RUN_ENS_POST = EMYSW.EMYSW.QINE.06.CS

# 5. Set the parameters for the second run.
OBS_SEGMENT_ID = FSMSW
OBS_DATA_TYPE = QINE

# 6. Perform the second run.
RUN_ENS_POST = FSMSW.FSMSW.QINE.06.CS
```

Figure 1.1. Example ENS_POST batch input file.

2.8 Output

The output generated by a run of ENS_POST is an error model adjusted ESP time series file. It will have the same complete name as the filename specified with the **RUN_ENS_POST** command, except that the last extension, which should usually be .CS, will now be the error model abbreviation specified by the **ERROR_MODEL** command. The example above would produce files by the names of EMYSW.EMYSW.QINE.06.ER2S and FSMSW.FSMSW.QINE.06.ER2S. Any output file should be visible from within ESPADP.

Appendix A. Error Models

ENS_POST operates by applying one error model to an ensemble of stream flow traces generated by ESP. As stated previously, there are, currently, four available error models. All of the error models perform the following basic steps when adjusting a stream flow trace:

1. Transform the stream flow values into normal space: This is accomplished through the following mathematical formula:

    $$z = Q^{-1}(F(x))$$

    where z is the transformed value, Q is the standard normal distribution function, F is the simulated cumulative distribution function for the current month and x is the value to transform.
2. Adjust the stream flow values in normal space: This is accomplished through the application of an error function.
3. Transform the adjusted stream flow values back into original space: This is accomplished through the following mathematical formula:

    $$y = G(Q^{-1}(z))$$

    where G is the observed cumulative distribution function for the current month and y is the adjusted flow value in original space.

The cumulative distribution functions are computed by the ENS_POST_CP and placed in files whose formats are described in Appendix C. Each error model requires certain parameters and these parameters are given in the parameter file. The parameter file is created by ENS_POST_CP and the format is given in Appendix B. The current error models are provided below with a general description of each.

1. ER1: This error model performs nothing for step 2. It just goes from step 1 directly to step 3.
2. ER2D: This error model applies a linear function resulting from an auto-regression of lag-1 for step 2. The regression includes the current simulated and the previous observed as predictors of the current observed. At time steps beyond the first, the previous adjusted value is used in place of the previous observed value in the regression. The result of applying the linear function is directly used in step 3.
3. ER2S: This is the same as ER2D, except that the result of the linear function is assumed to be the mean of a Normal distribution with a known variance. A random sample is drawn from this distribution and this value is used in step 3.

The last two error models are calibrated via a regression, which is performed by the program ENS_POST_CP. The last error model is the only error model that will add random noise to the adjusted flow time series in an attempt to account for uncertainties in the model itself.

Appendix B. Parameter File Format

The parameter file contains all of the parameters necessary for the ENS_POST to apply the Ensemble Post-Processor to a time series. It specifies the calibration time scale, the location of the simulated and observed CDF data, the maximum flow filter settings, monthly statistics and error model parameters. Each is described in the following sections.

B.1 Calibration Time Scale

The calibration time scale is the time scale for which the calibrations of the Ensemble Post-Processor were performed. It is equivalent to the time step given for a run of ENS_POST_CP. The format of the line specifying the time scale is

    calibration_time_scale = <num>

 where <num> is the time scale in hours. If this parameter is not present in the file, it is assumed to be 24 hours.

B.2 Simulated and Observed CDF Data

The next two lines of the parameter file specify the location of the cumulative distribution files (see Appendix C) relative to the subdirectory "error_model" of the directory corresponding to the apps-defaults token ens_input. The format of the two lines is

    <Sim or Obs>_CDF = <SEGID>/<SEGID>.<sim or obs>.cdf

where <Sim or Obs> is either "Sim" for the simulated CDF or "Obs" for the observed CDF; <SEGID> is the five letter segment id of the segment to which this parameter file applies; and <sim or obs> is the same as <Sim or Obs>, except in all lower case.

B.3 Monthly Statistics

The next four lines specify the monthly statistics of the simulated and observed data. The statistics given are mean and coefficient of variation. The format of the lines is

    <SIM or OBS>_MONTHLY_<STAT> = <value 1>,...,<value 12>

where <SIM or OBS> is either "SIM" for the simulated data or "OBS" for the observed data; <STAT> is either "MEAN" or "COEFF_VAR" depending on the statistic produced; and <value n> is the value of that statistic for the n-th month (1 = January, ..., 12 = December).

NOTE: These parameters are not necessary for a run of the post-processor.

B.4 Error Model Parameters

Currently, applying the Ensemble Post-Processor consists of applying one error model to the input time series. Each error model has a set of parameters necessary to apply that error model. The remaining lines of the parameter file specify these parameters. The format of each line is:

    <model name>_<parameter name>=<value 1>,...,<value 12>

where <model name> is the abbreviated name of the error model ("Model3S" for example); <parameter name> is the name of the parameter; and <value n> is the parameter value for the n-th month.

Appendix C. CDF File Format

The cumulative distribution function (cdf) file provides the monthly cumulative distribution of the historical observed and corresponding historical simulated data for a segment of interest. The format of

the file is as follows:

1. The first line specifies the total number of points contained in the file.
2. Every line thereafter contains a point from the cumulative distribution function, that is a flow followed by an empirical cumulative distribution value. There may be additional columns of data beyond these two values, but they will be ignored.

A file is produced by ENS_POST_CP for each month and each segment for which any error model is calibrated. Every error model requires these files.