

IVP Batch Program User's Manual For Verification

By

Henry Herr
Office of Hydrologic Development
National Weather Service

Table of Contents

1.0	<i>Overview</i>	3
1.1	<i>Using This Manual</i>	3
2.0	<i>Notation and Definitions</i>	3
3.0	<i>Execution</i>	3
4.0	<i>Directory Structure</i>	4
5.0	<i>Apps-Defaults Tokens</i>	4
6.0	<i>File Types</i>	5
7.0	<i>Batch File Format</i>	5
8.0	<i>Instructions</i>	7
9.0	<i>Batch Commands</i>	8
	ANALYSIS_INTERVAL = "<quantity> <unit>"	8
	BREAKDOWN_BY_LID = <ON/OFF>.....	8
	END_TIME = "<DATE>".....	8
	FCST_CAT = "<CAT1>,<CAT2>,...,<CATn>".....	9
	FCST_CAT_USED = <category>.....	9
	FCST_TS = "<TS1>,<TS2>,...,<TSn>".....	9
	GRAPH_TEMPLATE = <filename>.....	10
	LEADTIME_END = "<number of hours>" -or- "<quantity> <unit>".....	10
	LEADTIME_START = "<number of hours>" -or- "<quantity> <unit>".....	10
	LEADTIME_STEP = "<number of hours>" -or- "<quantity> <unit>".....	10
	OBS_CAT = "<CAT1>,<CAT2>,...,<CATn>".....	11
	OBS_CAT_USED = <category>.....	11
	PAIRS_FILE = "<filename>,<c/a>".....	11
	PE = "<PE1>,<PE2>,...,<PEn>".....	11
	PERSISTENCE = <ON/OFF>.....	12
	PRIMARY_STATS = "<batch stat>,<batch stat>,...".....	12
	PRIMARY_PLOT_TYPE = <type>.....	12
	RIVERRESPONSE = "<RESP1>,<RESP2>,...".....	12
	SECONDARY_STATS = "<batch stat>,<batch stat>,...".....	12
	SECONDARY_PLOT_TYPE = <type>.....	12
	START_TIME = "<DATE>".....	13
	XAXIS_VARIABLE = "<variable>".....	13
10.0	<i>Batch Actions</i>	14
	CALCSTATS = "<category>,<STAT1>,<STAT2>,...,<STATn>".....	14
	CLEAR_GROUPS = <anything>.....	14
	DEF_GRP = "<LID1>,<LID2>,...,<LIDn>".....	14
	DEF_LOC = "<LID1>,<LID2>,...,<LIDn>".....	15
	GEN_GRAPH = <image filename>[,<data filename>]".....	15
	NATLSTATS = "<LID1>,<LID2>,...,<LIDn>".....	15
	OUTPUT_FILE = "<filename>,<c/a>".....	16
11.0	<i>Special Tokens</i>	17
	@FILE = <filename>.....	17
	@+ = <line continuation>.....	17
12.0	<i>The Verification Data Gathering Algorithm</i>	18
13.0	<i>Examples</i>	20

1.0 Overview

The IVP Batch Program serves two functions:

- (1) constructing forecast-observed data pairs and
- (2) calculating verification statistics.

This manual provides instructions for the second function, calculating verification statistics using data stored in the vfpairs table of the archive database.

1.1 Using This Manual

This manual provides a description of the format of an IVP Batch Program input batch file. If producing raw statistics (numbers), it is recommended that the user review the commands in this manual and the examples prior to using the software. Also, read Section 8.0 for instructions on how to put together a batch input file. If producing graphics, it is recommended that the user run the IVP program (see the *Interactive Verification Program User's Manual*) to create the batch file. If editing is needed, this manual can be used to determine how to edit the batch file, and the IVP Batch Builder can be used to do the editing.

2.0 Notation and 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:

- *input token*: The <token> portion of a batch file input line. All tokens are displayed in this font.
- *token value*: The <value> portion of a batch file input line. All token values are displayed in this font and in quotes "".
- *batch command*: An input token that is used to specify a parameter. Its token value is stored by the batch file processor. A batch command does not result in any calculations or output (except logging output written to the terminal). Batch commands are displayed in **bold**.
- *batch action*: An input token that triggers a particular action, such as querying the archive database for pairs or performing calculation of verification statistics. The token value specifies parameters of that action. Batch actions are displayed in **bold**.
- *data pair*: A forecast-observed pair, defined in the vfpairs table of the archive database.
- *verification group*: A collection of locations, physical elements, etc., that are to be lumped together to produce one set of statistics.

3.0 Execution

Before executing the IVP Batch Program, be sure that the tables vfyruninfo, rivercrit, and location are all populated correctly for each location for which verification is to be done. The vfyruninfo table is populated using the Vfyruninfo Editor. The rivercrit table must be populated in order for the flood stage to be found, given by the field flood, if the flood stage is used in determining categories. The location

table must be populated in order for the rfc to be identified for a location, given by the field rfc. If no rfc is found, then the rfc is assumed to be "NONE".

To execute the Verification Batch Program, enter:

```
cd $(get_apps_defaults verify_dir)/scripts
ivpbatch [-c] <batch file name>
```

where <batch_file_name> is the name of the batch file. If the first letter of the name of the batch file is either '.' or '/', then the file name is assumed to be fully specified, relative to the current directory. Otherwise, it is assumed to be specified relative to the directory given by the apps-defaults token "vsys_input". Use the -c option only if executed within a cron.

4.0 Directory Structure

The following directory structure must be in place for the IVP Batch Program to execute properly:

```
$(vsys_dir)/bin/RELEASE/rfc.ob7.2.jar
$(vsys_dir)/bin/RELEASE/dbgen.jar
$(vsys_dir)/input/
$(vsys_dir)/files/$(LOGNAME)/templates/
$(vsys_dir)/output/$(LOGNAME)/
$(vsys_dir)/scripts/ivpbatch
```

\$(vsys_dir) refers to the value of the apps-defaults token vsys_dir, which points to the base directory, typically /rfc_arc/verify. \$(LOGNAME) refers to the user name; each directory above with this in it should be created for each user who is to use the software.

5.0 Apps-Defaults Tokens

The following apps-defaults tokens are used by the IVP:

- adb_name : <depends upon RFC>
- verify_dir : /rfc_arc/verify
- vsys_dir : \$(verify_dir)
- vsys_input : \$(vsys_dir)/input
- vsys_output : \$(vsys_dir)/output
- vsys_files : \$(vsys_dir)/files
- rax_pghost : ax <optional; if not present, it uses the PGHOST environment variable, instead>
- pguser : pguser <optional; if not present, it uses the PGUSER environment variable, instead>
- pgport : 5432 <optional; if not present, it uses th PGPORT environment variable, instead>

Each of the above directories must exist for the IVP Batch Program to run properly. It is also recommended that the following directory be created for each user:

```
$(vsys_input)/$LOGNAME
```

If this recommendation is followed, then the apps-defaults site file should override the setting of `vsys_input` as follows:

- `vsys_input` : `$(vsys_dir)/input/$(LOGNAME)`

All of these directories should be constructed prior to running IVP.

6.0 File Types

The following types of files are used within the IVP (recommended file extensions are given in parentheses; for the output image files, the extensions are required)

- **Batch Input Files (.bat)**: These files are input files to the IVP Batch Program, and can be loaded by the GUI in order to specify parameters of the data used and plots generated. These files are, by default, assumed to be in the `$(vsys_input)` directory.
- **Graph Template Files (.txt)**: These files specify properties of the charts to create, including labels, fonts, colors, sizes, axis limits, and others. These files are, by default, assumed to be in the `$(vsys_files)/$LOGNAME/templates` directory.
- **Output Image Files (.png, .jpg, .jpeg)**: These files are images of the charts generated by the IVP or IVP Batch Program. These files are, by default, assumed to be in the `$(vsys_output)/$LOGNAME` directory. Any such file **MUST** have one of the extensions listed.
- **Output Data Files (.dat)**: These files are ASCII format files that provide the data plotted to a chart in a tabular format. These files are, by default, assumed to be in the `$(vsys_output)/$LOGNAME` directory.
- **System Settings File**: This file is an ASCII format file used to specify defaults for the appearance of the IVP and its charts. The file must be in the directory `$(vsys_dir)/app-defaults` and have the name `IVP_SYSTEM_FILE.txt`. See Appendix A for more details.

7.0 Batch File Format

Each line of the batch file corresponds to a command or a parameter setting. All lines must be of the format:

```
<token> = <value>
```

with the following restrictions:

1. The token is not case sensitive.
2. Any number of spaces or tabs may be placed before or after the token and before or after the value.
3. A space, new-line (carriage return), tab, or pound (“#”) marks the end of a value or token.
4. Double quotes must be placed around the value if it is to contain tabs, spaces, or pounds, but the value may never contain a new-line. For example,

```
my_name = john doe
```

has a token of “my_name” and a value of “john”, whereas

```
my_name = "john doe"
```

has a token of "my_name" and a value of "john doe". If a new-line is encountered, it is treated as the closing double-quote.

5. The character '#', unless it is within double quotes, is used to indicate a comment. All characters after a '#' are ignored.
6. The equal sign ('=') must not be used as part of a value.

If a line is found which does not follow this format or specifies an unrecognized token, an error message will be generated. Blank lines are ignored.

8.0 *Instructions*

The following steps can be used to setup a batch input file to calculate verification statistics and output those statistics to an ASCII tabular output file:

1. Decide on the location and physical element pairs for which verification is to be done. For each location, if you wish to define forecast or observed value categories, then do so by using the `FCST_CAT` and `OBS_CAT`. Each location is then made available for verification via the `DEF_LOC` command. Categories can be set in either absolute terms or relative to the location's flood stage, as defined in the `rivercrit` table of the archive database.
2. Decide on the data time interval for which verification is to be done. Set the time frame in the batch file using the `START_TIME` and `END_TIME` commands.
3. If you wish to break down the overall time interval into subintervals, then determine the width of each subinterval and set it within the batch file using the `ANALYSIS_INTERVAL` command.
4. If you wish to only calculate statistics for forecasts with a lead time within a particular range of values, then determine the total lead time interval, and set it within the batch file using the `LEADTIME_START` and `LEADTIME_END` commands. If you wish to further break down the total lead time interval into subintervals, each with separately calculated statistics, then determine the width of each subinterval and set it within the batch file using the `LEADTIME_STEP` command.
5. If you wish to only calculate statistics for locations with specific river response times, then determine the response times you wish to use (slow, medium, or fast), and set it within the batch file via the `RIVERRESPONSE` command.
6. If you wish to limit the analysis to data values with particular physical elements or forecast type sources, then determine what physical elements and forecast type source you wish to use and set them in the batch file using the `PE` and `FCST_TS` commands.
7. Determine the verification groups you wish to construct and define the groups using the `DEF_GRP` command. A verification group defines a collection of locations for which one set of verification statistics are to be produced. A location can only be added to a group if it has a river response time within those response times given in Step 5, and if it has the same number of forecast categories and observed categories, defined in Step 1, as all other locations added to the group.
8. If you do not wish to use the default output file, then determine what output file name to use and open up the file by calling the batch action `OUTPUT_FILE`.
9. Determine the statistics you wish to produce and call the `CALCSTATS` action in the batch file passing in those statistics.
10. Use the created batch file as input to the IVP Batch Program. The desired output file will be generated.

If the user desires to produce graphics via the IVP Batch Program, the user should use the IVP to generate such a batch file.

9.0 Batch Commands

A batch command sets a parameter that is used by a batch action (see Section 8).

This section provides an alphabetical listing of all of the available batch commands. Commands associated with verification groups (see DEF_GRP action) are in **blue**; those associated with verification locations (see DEF_LOC action) are in **red**; those associated with both are in **purple**; and those associated with graphics (see GEN_GRAPH action) are in **green**. Acceptable values will be listed for each command, as well as the default if the command is not specified or if the command's token value is "<DEFAULT>". If the passed in token value is not acceptable, then the batch program will print an error message and stop. The following are batch commands:

ANALYSIS_INTERVAL = "<quantity> <unit>"

Description: Defines the analysis interval for the verification calculations. This interval breaks down the total [START_TIME, END_TIME] interval into subintervals, each of width equal to that amount given by the token's value. When statistics are calculated, they will be calculated independently for each interval. A value of "NONE" can be used to specify that no subintervals are to be created.

Acceptable Values: "<quantity> <unit>" or "MONTHLY". The <quantity> must be a positive integer and the <unit> must be either "WEEKS" ("WEEK" or "WK"), "DAYS" ("DAY" or "DY"), or "HOURS" ("HOUR" or "HR"). If "MONTHLY" is specified, then the overall interval will be broken down into months, with the first interval being from START_TIME to the end of its month and the last interval being from the first of the END_TIME month through the END_TIME.

Default Value: "NONE".

BREAKDOWN_BY_LID = <ON/OFF>

Description: If "ON", before computing statistics, the corresponding verification group(s) will be broken down so that each location (passed is as a location id in the value of the DEF_GRP action) is defined as a separate group. For example, to compute statistics independently for all locations available in the vfyruninfo table, first set this command to "ON" and then call DEF_GRP with a value of "ALL".

Acceptable Values: Either "ON" or "OFF".

Default Value: "OFF".

END_TIME = "<DATE>"

Description: Defines the end date/time for the pairing run. The end time can be absolute or can be relative to the current system time. Any data pair included in statistics calculation must have a valid time prior or equal to this date/time.

Acceptable Values: If the date is absolute, then it must be of one of these formats:

- "CCYY-MM-DD" (assumes time of 23:59:59 that day),
- "CCYY-MM-DD hh:mm:ss",
- "CCYY-MM-DD hh:mm:ss TZC",
- "MMDDCCYY:hh".

If the date is relative, then it must of the following format:

“* [$+ or -$ <quantity> <unit> <quantity> <unit> ...]”.

Everything in ‘[]’ is optional. The <quantity> must be a positive integer and the <unit> must be either “WEEKS” (“WEEK” or “WK”), “DAYS” (“DAY” or “DY”), or “HOURS” (“HOUR” or “HR”).

Default Value: “*” (the current system time).

FCST_CAT = "<CAT1>,<CAT2>,...,<CATn>"

Description: Defines categories used to break down data pairs by the forecast value. The categories are defined as follows: ($CAT1 \leq x < CAT2$), ..., ($CATn-1 \leq x \leq CATn$). Any data pair included in statistics calculation must have a forecast value within at least one of the defined categories. The categories can be location dependent if defined relative to the flood stage.

Acceptable Values: A list of categories or “NONE” to not categorize data. The list must be comma separated and, if you have spaces within the list, must be within double quotes. Each “<CATn>” value must be one of the following:

- a number (decimal or otherwise; for stage data, unit is assumed to be feet),
- “MIN” to denote no lower bound,
- “MAX” to denote no upper bound,
- “<scalar>*<stage>”.

In the last case, the scalar is a positive integer value, and the stage is either “AS”, “FS”, “ModFS”, “MajFS”, or “RS”. These correspond to stages defined in the rivercrit table: action stage, flood stage, moderate flood stage, major flood stage, and record stage. If the desired stage cannot be found for a location, a message will be generated stating that no flood stage was found and the batch processor will print an error message and stop. The list will be sorted into ascending order prior to use. See Section 10 for examples.

Default Value: “NONE”, which is equivalent to “MIN, MAX”.

FCST_CAT_USED = <category>

Description: Sets the forecast category to use. See Section 13.5 of the *Interactive Verification Program User's Manual* for a description of this command. Either this command or OBS_CAT_USED must be something other than “NONE”. If this command is set to “NONE” and OBS_CAT_USED is already “NONE”, then OBS_CAT_USED will be set to “ALL”. If this is set to a value other than “NONE”, then OBS_CAT_USED will be set to “NONE”.

Acceptable Values: Must be “NONE” (equivalent to “Do Not Use” in IVP), “ALL” (“All Categories Combined”/“Use Only Category” in IVP), or “CAT#” (“Category #” in IVP) where ‘#’ is the category number (the categories are sorted into ascending order and numbered 1, 2,...).

Default Value: “NONE”

FCST_TS = "<TS1>,<TS2>,...,<TSn>"

Description: Defines a list of forecast type sources. Any data pair included in statistics calculation must have a forecast type source in this list.

Acceptable Values: A list of valid forecast type sources or “ALL” to allow for any valid type source. The list must be comma separated and, if there are spaces within the list, must be within double quotes. If

“ALL” is specified, then a list of all forecast type sources is generated from the vfyruninfo table, and that list is used to query the database for data pairs.

Default Value: “ALL”.

GRAPH_TEMPLATE = <filename>

Description: Defines the name of a template file that contains chart properties for the chart to create (via GEN_GRAPH).

Acceptable Values: A filename of a file that can be read on the system, or “NONE”. If the filename does not begin with a ‘.’ or ‘/’, then it is assumed to be located within the directory corresponding to \$(vsys_files)/\$(LOGNAME)/templates, where \$(vsys_files) is the directory corresponding to apps-defaults token vsys_files, and \$(LOGNAME) is the subdirectory corresponding to environment variable LOGNAME. If “NONE” is specified, then default chart properties are used.

Default Value: “NONE”.

NOTE: Though the file is in ASCII format, it is highly recommended that only the IVP **Chart Property Manager** window be used to edit the properties contained in a template file. Hence, the parameters of the template file will not be discussed in any, User's Manual.

LEADTIME_END = “<number of hours>” -or- “<quantity> <unit>”

Description: Defines the largest lead time to be used in verification. Any data pair included in statistics calculation must have a (validtime – basistime) smaller than or equal to this upper bound. A value of “NONE” can be used to specify no upper bound.

Acceptable Values: “NONE”, a positive integer, or see ANALYSIS_INTERVAL if “<quantity> <unit>” is used. The value of this token, in hours, must be larger than LEADTIME_START.

Default Value: “NONE”.

LEADTIME_START = “<number of hours>” -or- “<quantity> <unit>”

Description: Defines the smallest lead time to be used in verification. Any data pair included in statistics calculation must have a (validtime – basistime) greater than this lower bound. A value of “NONE” can be used to specify no lower bound (this is equivalent to a LEADTIME_START of “0”).

Acceptable Values: “NONE”, zero, a positive integer, or see ANALYSIS_INTERVAL if “<quantity> <unit>” is used.

Default Value: “NONE”.

LEADTIME_STEP = “<number of hours>” -or- “<quantity> <unit>”

Description: Defines the lead time interval for the verification calculations. This interval breaks down the total (LEADTIME_START, LEADTIME_END] interval into subintervals, each with a width equal to the token's value. When statistics are calculated, they will be calculated independently for each lead time interval. Note that each interval is open at the lower end and close at the upper end. For example, if LEADTIME_STEP is 6 and LEADTIME_START is 0, then the first interval will be (0,6], the next will be (6, 12], and so on. A value of “NONE” can be used to specify that no subintervals are to be created.

Acceptable Values: “NONE”, a positive integer, or see ANALYSIS_INTERVAL if “<quantity> <unit>” is used.

Default Value: “NONE”.

OBS_CAT = "<CAT1>,<CAT2>,...,<CATn>"

Description: Defines categories used to break down data pairs by the observed value. The categories are defined as follows: ($CAT1 \leq x < CAT2$), ..., ($CAT_{n-1} \leq x \leq CAT_n$). Any data pair included in statistics calculation must have an observed value within at least one of the defined categories. The categories can be location dependent if defined relative to the flood stage.

Acceptable Values: See FCST_CAT command above, except that the categories bound the observed value, not the forecast value.

Default Value: "NONE".

OBS_CAT_USED = <category>

Description: Sets the observed category to use. See Section 13.5 of the *Interactive Verification Program User's Manual* for a description of this command. Either this command or FCST_CAT_USED must be something other than "NONE". If this command is set to "NONE" and FCST_CAT_USED is already "NONE", then FCST_CAT_USED will be set to "ALL". If this is set to a value other than "NONE", then FCST_CAT_USED will be set to "NONE".

Acceptable Values: Must be "NONE" (equivalent to "Do Not Use" in IVP), "ALL" ("All Categories Combined"/"Use Only Category" in IVP), or "CAT#" ("Category #" in IVP) where '#' is the category number (the categories are sorted into ascending order and numbered 1, 2,...).

Default Value: "NONE"

PAIRS_FILE = "<filename>,<c/a>"

Description: Defines the name of the pairs file to use and whether to open it for creation or append.

Acceptable Values: Either "NONE" or a filename, followed by a comma, and then either 'c' for create or 'a' for append. If the filename does not have either a '.' or a '/' at the first character, then it will be assumed that the file is to be placed relative to the directory given by the apps-default token "vsys_output". If "NONE" is given, then the pairs data will not be output to any file. See the DEF_GRP action (Section 8) for more details.

Default Value: "NONE".

NOTE: If you wish to view the pairs data within the IVP graphical user interface, then you will need to create a pairs_file first, and the pairs file absolutely must contain the string ".pairs" in its name. It is this pairs file that the IVP GUI reads as a source of data.

PE = "<PE1>,<PE2>,...,<PEn>"

Description: Defines a list of physical elements. When a location is defined via a DEF_LOC action, this restricts the locations defined to be those which have a specified physical element from within this list. When a group is defined via a DEF_GRP action, this restricts the locations added to the group to be those with a specified physical element from within the list.

Acceptable Values: A list of valid physical elements or "ALL" to allow for any valid physical element. The list must be comma separated and, if you have spaces within the list, must be within double quotes. If "ALL" is specified, then a list of all physical elements is generated from the vfyruninfo table, and that list is used to query the database for data pairs.

Default Value: "ALL".

PERSISTENCE = <ON/OFF>

Description: Determines if persistence forecast pairs are to be used in computing statistics. If set to ON and if FCST_TS is set to ALL (its default value), then pairs with a persistence forecast type source will be included in verification. If set to OFF, then the only way to include persistence forecasts in verification is to explicitly list the persistence forecast type source ("FR") in the value of a FCST_TS command.

Acceptable Values: Either "ON" or "OFF".

Default Value: "ON".

PRIMARY_STATS = "<batch stat>,<batch stat>,..."

Description: Sets the primary statistics, or those statistics displayed against the left-hand y-axis of the generated plot.

Acceptable Values: Each <batch stat> must be the short hand notation for a batch statistic, as described in Section 14.0, on the **IVP Statistic Chooser Manager**, in the *Interactive Verification Program User's Manual*. All of the statistics must have the same scale, which means they must be in the same group within the **IVP Statistic Chooser Manager** of the IVP.

Default Value: No default value. This command must have a value.

PRIMARY_PLOT_TYPE = <type>

Description: Sets the plot type to use for the primary statistics in the generated plot.

Acceptable Values: Must be "BAR", "LINE", or "SCATTER".

Default Value: "SCATTER"

RIVERRESPONSE = "<RESP1>,<RESP2>,..."

Description: Defines a list of river response times. The response time is stored in the vfyruninfo table of the archive database as the resptime field. Any location included in statistics calculation must have a response time in this list. If no resptime field can be found for a given location in the vfyruninfo table, then that location will be assigned a response time of "NONE".

Acceptable Values: A list of valid response times ("SLOW", "MEDIUM", "FAST") or "ALL" to allow for any response time, including "NONE". The list must be comma separated and, if you have spaces within the list, must be within double quotes.

Default Value: "ALL".

SECONDARY_STATS = "<batch stat>,<batch stat>,..."

Description: Sets the secondary statistics, or those statistics displayed against the right-hand y-axis of the generated plot.

Acceptable Values: "NONE" or Each <batch stat> must be the short hand notation for a batch statistic, as described in Section 14.0, on the **IVP Statistic Chooser Manager**, in the *Interactive Verification Program User's Manual*. All of the statistics must have the same scale, which means they must be in the same group within the **IVP Statistic Chooser Manager** of the IVP.

Default Value: "NONE"

SECONDARY_PLOT_TYPE = <type>

Description: Sets the plot type to use for the secondary statistics in the generated plot.

Acceptable Values: Must be "BAR", "LINE", or "SCATTER".

Default Value: "SCATTER"

START_TIME = "<DATE>"

Description: Defines the start date/time for the pairing run. The start date/time can be absolute or can be relative to the current system time. Any data pair included in statistics calculation must have a valid time after or equal to this date/time.

Acceptable Values: See END_TIME above, except that for the format of "CCYY-MM-DD" a time of 00:00:00 is assumed.

Default Value: "* - 14 DAYS" (two weeks prior to current system time).

XAXIS_VARIABLE = "<variable>"

Description: Defines the variable to display along the x-axis, against which the statistics are to be plotted in the generated plot.

Acceptable Values: Must be "LOCATION", "ANALYSIS INTERVAL", "LEADTIME INTERVAL", "OBSERVED CATEGORY", or "FORECAST CATEGORY".

Default Value: "LOCATION"

10.0 Batch Actions

Actions instruct the verification program to do something, such as open a file or calculate statistics. The nature of what is done depends on the action given. Acceptable values will be listed for each token value. If the value is not acceptable, then the batch program will print an error message and stop.

The following are valid actions within the verification system:

CALCSTATS = "<category>,<STAT1>,<STAT2>,...,<STATn>"

Description: This action causes statistics to be generated given the current command settings provided prior to this line in the batch file. Statistics will be produced for every verification group. All output is sent to the output file specified by the latest execution of the OUTPUT_FILE action.

Acceptable Values:

- <category>: Specifies if categories are to be constructed relative to observed values, forecast values, or both observed and then forecast. The value must be either "OBS" for observed categories, "FCST" for forecast categories, or "BOTH" for both categories.
- <STATn>: Specifies a statistic to produce. Values for <STATn> can be one of the following:
 - ERRORS: All of the statistics root mean square error, maximum error, mean error, and mean absolute error.
 - CATSTATS: All of the statistics probability of detection, traditional false alarm rate, hydrologic false alarm rate, under forecast rate, and over forecast rate.
 - QUANTILES: The minimum, 25% quantile, median, 75% quantile, and maximum values for the non-category variable in each category.

Default Value: Does not apply. The value must be acceptable.

NOTE: If the output file has not been opened via the OUTPUT_FILE action below when CALCSTATS is called, then the default name will be used, and it will be created.

CLEAR_GROUPS = <anything>

Description: Clears all of the groups that have been added up until this point in the batch file.

Acceptable Values: The value is ignored.

Default Value: Does not apply. The value is ignored.

DEF_GRP = "<LID1>,<LID2>,...,<LIDn>"

Description: Defines a group as a collection of location ids, each one of which must have been defined at least once via a DEF_LOC action previously within the batch file. In addition to a list of locations, associated with each group is a list of physical elements, forecast type sources, start time, end time, analysis interval, lead time start, lead time end, lead time step, pairs output file, and a flag specifying if the group is to be broken down by location prior to use. These are given by the commands PE, FCST_TS, START_TIME, END_TIME, ANALYSIS_INTERVAL, LEADTIME_START, LEADTIME_END, LEADTIME_STEP, PAIRS_FILE, and BREAKDOWN_BY_LID, respectively. If the pairs output file is defined as "NONE", then no data pairs will be output to a file. However, if it is a valid file, then all of the data pairs for this group will be output to that file.

Acceptable Values:

- <LIDn>: A defined location (via DEF_LOC, described below) or "ALL" to use every location currently defined via DEF_LOC. Note that:
 - All locations must be defined for at least one physical element via the DEF_LOC command.
 - All locations within a group must have an identical number of forecast categories and an identical number of observed categories.
 - If any location is specified does not have a response time that is one of the response times provided in the latest execution of the RIVERRESPONSE command, then it will be left out of the group.

Default Value: None.

DEF_LOC = "<LID1>,<LID2>,...,<LIDn>"

Description: Defines the locations given in the list using parameters as set via commands in the batch file prior to this action. This action uses the PE command to determine the physical elements for which to define the given locations. Associated with each location and physical element pair are forecast and observed categories, given by commands FCST_CAT and OBS_CAT, respectively.

Acceptable Values:

- <LIDn>: A location to define for which a record must exist within the vfyruninfo table, or "ALL" to define all locations given within vfyruninfo to have the categories.

Default Value: None.

GEN_GRAPH = <image filename>[,<data filename>]"

Description: Generates a graphic, using the values of commands PRIMARY_STATS, SECONDARY_STATS, PRIMARY_PLOT_TYPE, SECONDARY_PLOT_TYPE, XAXIS_VARIABLE, FCST_CAT_USED, OBS_CAT_USED to define the parameters of a plot, exactly as the corresponding fields in the **Verification Plot Definition Manager** defines the plot for the IVP (see Section 13 of the *Interactive Verification Program User's Manual*). The value of command GRAPH_TEMPLATE defines a template file to be read which contains properties for the displayed chart. These properties are those defined in the **Chart Property Manager** of the IVP. If more than one verification group has been defined via a DEF_GRP action, then only the first group will be used to generate a graphic.

Acceptable Values:

- <image filename>: The image file to create. The filename must have a ".jpeg", ".jpg", or ".png" extension, so that the IVP Batch Program knows which kind of image file to create.
- <data filename>: Optional. The name of the data file to create. This file is an ASCII file, equivalent to clicking on "Save to File" in the **IVP Statistics Data Viewer** of the IVP (see Section 16 of the *Interactive Verification Program User's Manual*).

Default Value: None. The <image filename> must be specified, though the <data filename> is optional.

NATLSTATS = "<LID1>,<LID2>,...,<LIDn>"

Description: This action causes national statistics to be generated for each of the passed in locations. All previously defined commands are used as with CALCSTATS, except that the following are assumed: ANALYSIS_INTERVAL = NONE, FCST_CAT = NONE, LEADTIME_END = 72, LEADTIME_START = 0, LEADTIME_STEP = 6, OBS_CAT = "*0.0,*1.0,*10.0", PERSISTENCE = OFF. It is assumed that every passed in location is to have its statistics calculated independently. Therefore, any previously defined verification locations and groups are ignored. Locations can still be restricted by using the PE command, and forecast type sources can be

restricted by using the FCST_TS command. The OUTPUT_FILE is ignored: the output statistics produced are organized into files by location RFC, leadtime interval, and category. The names of the output files are:

```
"<rfc>_<ccyyymm>_<above/below>.stat_<lead time interval end>hr_tab"
```

The file is placed in the directory given by the value of the apps-defaults token "vsys_output". <rfc> is the lower case five letter rfc abbreviation; <ccyyymm> is the four digit year and two digit month of the date specified by the batch command START_TIME; <above/below> is either "above" for data in the upper category or "below" for data in the lower category; and <lead time interval end> is the upper bound of the lead time interval of the statistics contained in the file.

Acceptable Values:

- <LIDn>: Specifies locations for which statistics are to be produced. Each location id must be present within the vfyruninfo database table.

Default Value: None.

OUTPUT_FILE = "<filename>,<c/a>"

Description: Opens up a file to store statistics. All statistics output, other than that generated by a NATLSTATS action, are written to this file. If an output file was opened previously, it will be close.

Acceptable Values:

- <filename>: The name of the file to open or "<DEFAULT>" to use the default name. If the first character is not a '/' or '.', then the file will be opened relative to the directory pointed to by apps-defaults token "vsys_output".
- <c/a>: Either 'c' or 'a'. If 'c', then the file specified by the <filename> will be created or overwritten. If 'a', then the file specified by the <filename> will be opened for appending.

Defaults: None. However, if an OUTPUT_FILE is not called when CALCSTATS is called, all output will be sent to an output file as if the action "OUTPUT_FILE = vfyoutput.txt,c" was called.

11.0 Special Tokens

A special token is a batch language token that is not related to specifying parameters or performing actions related to the software program.

This section provides an alphabetical listing of all special tokens. The following are special tokens:

@FILE = <filename>

Description: Forces batch processor to process the file specified by <filename>. The file is treated just as if its contents were included within the original batch file.

Acceptable Values: Any valid file that will not force the batch processor to enter into an infinite, recursive loop (i.e. if file 1 references file 2, then file 2 can never reference file 1).

Default Value: Does not apply. The file name must be valid.

@+ = <line continuation>

Description: Line continuation token. The batch processor will take the value of this line and append it to the value on the previous line.

Acceptable Values: Depends upon previous lines batch command or action.

Default Value: Depends upon previous lines batch command or action.

12.0 The Verification Data Gathering Algorithm

The algorithm used to collect verification data pairs from the vfy pairs table of the archive database is as follows:

1. Every time the action `DEF_LOC` is called, a collection of locations are defined for the specified location ids and user specified list of pes (via `PE` command). Associated with each defined location in the collection are the following:
 - List of forecast value categories given by the `FCST_CAT` command. This is stored in its unprocessed (comma separated list) form.
 - List of observed value categories given by the `OBS_CAT` command. This is stored in its unprocessed (comma separated list) form.
 - Other information, including RFC, response time, and flood stage, each of which is loaded from the archive database (tables `location`, `vfyruninfo`, and `rivercrit`, respectively).

2. Every time the action `DEF_GRP` is called, a verification group is defined. Associated with each defined group are the following:
 - List of locations in the group.
 - List of acceptable physical elements (`PE` command).
 - List of acceptable forecast type sources (`FCST_TS` command).
 - Analysis window start and end times (`START_TIME`, `END_TIME` commands) and analysis interval (`ANALYSIS_INTERVAL` commands).
 - Lead time interval start and end (`LEADTIME_START`, `LEADTIME_END` commands), and lead time step (`LEADTIME_STEP` commands).
 - Flag specifying if group is to be broken down by location prior to use.

When `DEF_GRP` is executed, each location's river response time is checked against the list of river responses provided by the command `RIVERRESPONSE`. If it is not found, then it will not be added to the group.

3. Every time the action `OUTPUT_FILE` is called, a file is opened based upon the parameters passed in for the action, and is left open until the next call to `OUTPUT_FILE` or until the program ends.

4. When the `CALCSTATS` algorithm is called, the first step performed is, if the any groups are to be broken down by location prior to use (see `BREAKDOWN_BY_LID` command), then each such group is redefined so that one group exists per location. Then, for each group in the list of groups and for each analysis interval in the list of analysis intervals for that group, the following is done:
 - a. All vfy pairs records are loaded that have a location, physical element, and forecast type source for the current group, a forecast validtime within the current analysis interval, and a forecast (validtime – basistime) within the overall lead time interval (from `LEADTIME_START` to `LEADTIME_END`) for the current group.
 - b. For each lead time subinterval (based on `LEADTIME_STEP`), a subset of the current vfy pairs records is created.

- c. If the first argument of the CALCSTATS command is either "BOTH" or "FCST", then for each subset from Step 4b and forecast category, the following is done:
 - i. All vfpairs records are extracted from the list gathered in Step 4b for which the forecast value is in the current category.
 - ii. All statistics are produced for the current records.
 - iii. All statistics are output to the output file in the appropriate format.
- d. If the first argument of the CALCSTATS command is either "BOTH" or "OBS", for each observed category, Steps 4c-i, 4c-ii, and 4c-iii, are each performed, except that in step 4c-i, the observed value is checked in place of the forecast value.

13.0 Examples

The following examples illustrate how to construct batch files to accomplish particular goals.

Example 1: A batch file to produce verification statistics for October 2003 for the national verification project. The locations used have lids of AAAAA,BBBBB,CCCCC, and DDDDD. The files to create will have prefix "stats_OCT_2003" and will be in the directory corresponding to apps-defaults token "vsys_output".

```

START_TIME = "20031001 00:00:00"
END_TIME   = "20031031 23:59:59"

# Do the statistics...
NATLSTATS  = "AAAAA,BBBBB,CCCCC,DDDDD"

```

Example 2: A batch file to produce verification statistics for October 2003 and for all locations defined within the vfyruninfo table with slow or medium response times. Only overall statistics are needed; categories are not needed. The statistics to produce are the error statistics ("ERRORS") and will be written to the default file. Persistence and normal forecasts are included in the verification.

```

DEF_LOC      = ALL

START_TIME   = "20031001 00:00:00"
END_TIME     = "20031031 23:59:59"
RIVERRESPONSE = "SLOW, MEDIUM"
DEF_GRP      = ALL

# Do the statistics...
# For overall statistics, I can pass any of the three
# acceptable values as the first argument.
CALCSTATS    = obs,ERRORS

```

Example 3: As example 2, except that for location AAAAA, which is a slow response time river, create two categories for the observed value: ($2.0 < x \leq 15.0$), and (above 15.0). For all other locations, use the categories below flood stage and above flood stage. Also, add the categorical statistics.

```

# First, define all locations.
OBS_CAT      = "MIN, *1.0 ,MAX"
DEF_LOC      = ALL

# Override the categories for lid AAAAA by calling
# DEF_LOC again for it only.
OBS_CAT      = 2.0,15.0,MAX
DEF_LOC      = AAAAA

# Now define the general group parameters.

```

```

START_TIME      = "20031001 00:00:00"
END_TIME        = "20031031 23:59:59"
RIVERRESPONSE  = "SLOW, MEDIUM"

# Do the statistics...
DEF_GRP         = ALL
CALCSTATS      = "obs, errors, catstats"

```

Example 4: Same as example 3, except produce statistics for lead times between 0 and 72 hours and at 24 hour time steps and only for locations AAAAA, BBBB, and CCCC. Also, send the output to a file with name "outfile.txt" relative to the apps-defaults token "vsys_output". Finally, do not include persistence forecasts in verification. The file should be appended to as it may already contain previously produced statistics.

```

# First, define the BBBB and CCCC locations.
OBS_CAT        = "MIN, *1.0 ,MAX"
DEF_LOC        = BBBB,CCCC

# Now define the AAAAA location.
OBS_CAT        = 2.0,15.0,MAX
DEF_LOC        = AAAAA

# Now define the general group parameters.
START_TIME     = "20031001 00:00:00"
END_TIME       = "20031031 23:59:59"
RIVERRESPONSE  = "SLOW, MEDIUM"
LEADTIME_START = 0
LEADTIME_END   = 72
LEADTIME_STEP  = 24
PERSISTENCE    = OFF

# Do the statistics...
DEF_GRP        = AAAAA,BBBB,CCCC
OUTPUT_FILE    = "outfile.txt,a"
CALCSTATS     = "obs, error, catstats"

```

Example 5: Same as example 4, except break apart the persistence forecasts, which have a type source of "FR", from the non-persistence forecasts. This will yield two groups.

```

# First, define the BBBB and CCCC locations.
OBS_CAT        = "MIN, *1.0 ,MAX"
DEF_LOC        = BBBB,CCCC

# Now define the AAAAA location.
OBS_CAT        = 2.0,15.0,MAX
DEF_LOC        = AAAAA

# Now define the general group parameters.
START_TIME     = "20031001 00:00:00"
END_TIME       = "20031031 23:59:59"

```

```

RIVERRESPONSE = "SLOW, MEDIUM"
LEADTIME_START = 0
LEADTIME_END   = 72
LEADTIME_STEP  = 24

# Define the two groups. The first includes all non-persistence
# forecasts.
PERSISTENCE    = off
DEF_GRP        = AAAAA,BBBBB,CCCCC

# The second group includes only persistence forecasts.
FCST_TS        = "FR"
DEF_GRP        = AAAAA,BBBBB,CCCCC

# Do the statistics...
OUTPUT_FILE    = "outfile.txt,a"
CALCSTATS     = "obs, error, catstats"

```

Example 6: Batch file used to generate a graphic in which the error statistics RMSE, MaxErr, MAE, and ME, and the sample size, are plotted against the location id (on the x-axis). Six locations are involved and all data is analyzed for the time period 9-1-2003 through 10-31-2003 and lead times between 0 and 5 days. No analysis subintervals and no lead time subintervals are used. Only data with a physical element of "HG" and forecast type source of either "FE" or "FF" is analyzed. No restriction is placed on the observed or forecast value (via the category defining commands). Both an image and a data file are to be constructed, and a template file is to be used.

Note that the second defined group, with locations "XXXXX,YYYYY,ZZZZZ" will be ignored when generating the graph, since it is the second defined group. Also note the line continuation used to define the locations.

```

#Define locations.
PE = HG
DEF_LOC = AAAAA,BBBBB,CCCCC,DDDDD,EEEEEE,FFFFFF,
@+      = XXXXX,YYYYY,ZZZZZ
START_TIME = "2003-09-01 00:00:00"
END_TIME   = "2003-10-31 23:59:59"
LEADTIME_START = 0hours
LEADTIME_END   = 5days
FCST_TS      = FE,FF

#Define the groups, but make sure that when processed the group will
#be broken down by location in order to acquire statistics for each
#location independent of the others. This allows for the graphic to
#display stats per location.
BREAKDOWN_BY_LID = ON
DEF_GRP = " AAAAA,BBBBB,CCCCC,DDDDD,EEEEEE,FFFFFF "
DEF_GRP = "XXXXX,YYYYY,ZZZZZ"

#Define the graphic.
PRIMARY_STATS = RMSE,MAXERR,MAE,ME
PRIMARY_PLOT_TYPE = SCATTER
SECONDARY_STATS = "NUM SAMPLES"
SECONDARY_PLOT_TYPE = BAR

```

```

XAXIS_VARIABLE = LOCATION
FCST_CAT_USED = ALL
OBS_CAT_USED = NONE

#Generate the graphic using the desired template file.
GRAPH_TEMPLATE = "example6.txt"
GEN_GRAPH = "example6.png,example6.dat"

```

Example 7: Produce the same graphic, but this time the x-axis variable is the lead time subinterval. Also, only the location XXXXX is used and the information necessary to define the group is in the batch file associated with Example 6, above. That file is called example6.bat.

```

#Load the group settings from the example 6 batch file.
@File = example6.bat
LEADTIME_STEP="1 days"

#Define the one group.
BREAKDOWN_BY_LID = ON
DEF_GRP = XXXXX

#Define the graphic.
PRIMARY_STATS = RMSE,MAXERR,MAE,ME
PRIMARY_PLOT_TYPE = SCATTER
SECONDARY_STATS = "NUM SAMPLES"
SECONDARY_PLOT_TYPE = BAR
XAXIS_VARIABLE = "LEADTIME INTERVAL"
FCST_CAT_USED = ALL
OBS_CAT_USED = NONE

#Generate the graphic using the desired template file.
GRAPH_TEMPLATE = "example6.txt"
GEN_GRAPH = "example7.png, example7.dat"

```
