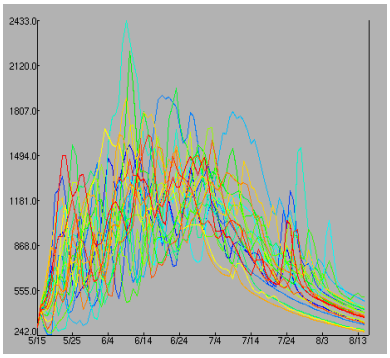
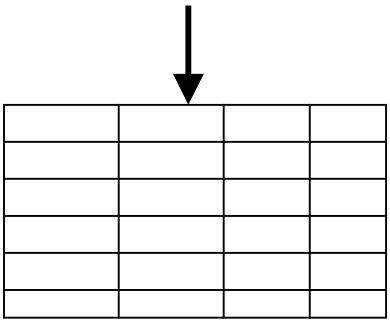
A Method for Storing ESP Forecasts In an INFORMIX Table



ESP



Database

David G. Brandon
Hydrologist In Charge
Colorado Basin River Forecast Center

Jun 26, 2001

A Method for Storing ESP Forecasts In an INFORMIX Table

Preface

This paper presents a method to store ESP forecasts into a RDBMS. It is offered to present a few concrete ideas and 'a solution' to 'a problem'. Hopefully, discussion will follow and additional ideas will surface that can better help define, refine and solve this problem.

1.0 *Introduction*

Ensemble Streamflow Prediction (ESP) is the foundation and linchpin of the AHPS. As AHPS projects expand in the NWS, the use of ESP increases at the RFCs and WFOs. It is evident that a procedure would be useful that allows a user to organize and retain the essential elements of ESP forecasts.

It would be desirable to utilize an existing NWS approved data/forecast exchange format, for example SHEF. A natural place for storing forecasts would seem to be the relational DBMS, INFORMIX.

The CBRFC has developed a method that allows ESP forecasts to be stored in an INFORMIX Table. The INFORMIX table structure accepts all of the unique attributes of an ESP forecast. The SHEF PEDTSEP was used to identify many of the attributes of the ESP forecast. Software has been developed that converts ESPADP output into an interim format. This format is passed to a parser/poster which stores the ESP forecasts into the INFORMIX table.

2.0 *ESP Forecast Review*

An ESP forecast is a strange 'animal'. Unlike a deterministic forecast, which is a single value, an ESP forecast consists of many values based upon a user-selected probability distribution. They can take the form of sums, means, instantaneous values, number of days and an extremum for any time window. They can result from several model types. They can be based on five probability distributions, at least 5 trace weighting schemes, different carryover groups, and different starting dates within the carryover group. Forecasts are also derived from time series that are of a certain 'data type' and 'time step'. Finally, the historical data that are analyzed have a specific beginning and ending date.

ESP forecasts can be classified into two categories. One category describes an ESP forecast single element that results from andanalysis window'. An example would be the instantaneous maximum flow that occurs during the next 3 months. The second category describes a time series of values for the 'analysis window'. An example of this would be the mean daily flow for each day for the next two weeks. Procedures have been developed to handle both of these categories.

The following is a list of the unique attributes of an ESP forecast. A detailed description of each attribute follows.

Basin/Segment ID
Type of Forecast
Model
Data Value
Probability of the Data Value
Date/Time of Analysis Window

Date of Creation
Probability Distribution
Weighting Scheme
Date/Time of Carryover
Carryover Group ID
Data Type of the Time Series
Time Step of the Time Series
Beginning Historical Year
Ending Historical Year

The following are additional information that 'may' be needed for certain cases.

Date/Time of Extremum in the Analysis Window
(for certain forecast types)
ND_Value
(If forecast type is NDTO(1), NDTO(2), NDIS(1), or NDIS(2))
Units of ND_Value
(If forecast type is NDTO(1), NDTO(2), NDIS(1), or NDIS(2))

Basin/Segment ID:

An eight character identifier which uniquely defines the NWSRFS segment.

Forecast Types:

MXND	**Maximum Mean Daily & Days to Max**
MNMD	**Minimum Mean Daily & Days to Min**
MD	**Mean Daily**
SUM	**Cumulative Sum (Volume)**
MXIN	**Maximum Instantaneous & Days to Max**
MNIN	**Minimum Instantaneous & Days to Min**
NDTO(1)	**Number of Days to Get Above a Value**
NDTO(2)	**Number of Days to Get Below a Value**
NDIS(1)	**Number of Days Greater than a Value**
NDIS(2)	**Number of Days Less Than a Value**

Model Sources:

Historical Observed
Historical Simulated
Conditional Simulated

Extremum:

A notation of whether the value is a maximum or minimum.

Date/Time of Analysis Window:

A description of the 'window' or time period, which the ESP forecast type is being analyzed. This may be described as a starting date/time with a duration, and ending date/time with a duration, or a beginning and ending date/time of the 'window'.

Date/Time of Creation:

The date and time the ESP forecast was created.

Probability distributions:

Empirical
Normal
Log Normal
Wakeby
Weibull

Weighting Schemes:

Equal (Climatology)
CPC Pre Adjustment Technique
CPC Post Adjustment Technique
Alaska Technique
Your Customize Technique(s)

Carryover Date/Time:

The date and time of the NWSRFS carryover group used when running the ESP program.

Carryover_Group ID:

An alphanumeric field representing the carryover group used in the forecast.

Data Type of the Time Series:

The NWSRFS data type used when the ESP run was made.

Time Step of the Time Series:

The time step of the NWSRFS data type used when the ESP run was made.

Beginning Historical Year:

The beginning year of the historical data that are being analyzed when the ESP run was made.

Ending Historical Year:

The ending year of the historical data that are being analyzed when the ESP run was made.

Date/Time of Extremum:

A date and time of the extremum predicted in the analysis window if the value is an extremum.

ND_Value:

The value specified if the forecast type is NDTO(1), NDTO(2), NDIS(1), or NDIS(2).

Units of ND_Value:

The units of the nd_value:
FT = feet
CFS = flow

Data Value:

The value(s) of the forecast associated with the probability.

3.0 *Translation*

The current rules of SHEF do not allow the capability to describe all of the unique attributes of an ESP forecast. Some of the attributes can match to the current PEDTSEP definitions. Other attributes cannot. A solution was needed to address this dilemma.

An initial approach used a new concept that utilized the 'retained comment' field that is part of SHEF. Special codes or tokens and associated values were coded in the retained comment field according to defined rules. This 'retained comment' field could be passed and decoded by a poster that is designed to post to the target ESP forecast table. After several attempts it became apparent that this approach produced SHEF messages where many of the attributes were contained in the retained comment field. Furthermore, it was very cumbersome to produce time series of values representing differing probability values.

A solution was developed that utilizes a CSV, or comma separated value format to encode and transmit the attributes and values of the esp forecasts. This format incorporates the SHEF PEDTSEP code to describe some of the attributes once they are posted to the database table.

3.1 *SHEF TRANSLATION FOR SOME OF THE ATTRIBUTES*

SHEF Physical Element: PEDTSEP

QC Runoff Volume
QR Discharge, River

SHEF Duration: PEDTSEP

V Variable

SHEF Type/Source: PEDTSEP

PE	Process #5	(Historical Simulated)
HE	Historical with #5 (new code)	(Historical Observed)
FU	Forecast Unadjusted Model 1	(Conditional Simulated)

SHEF Extremum: PEDTSEP

K	Minimum of Year calendar (assumed period)
U	Maximum of Year Calendar (assumed period)
Z	Default

SHEF Probability: PEDTSEP

A	.002
B	.004
C	.01
D	.02
E	.04
F	.05
1	.1
2	.2
G	.25
3	.3
4	.4
5	.5
6	.6
7	.7
H	.75
8	.8
9	.9
T	.95
U	.96
V	.98
W	.99
X	.996
Y	.998

3.2 *Translation For The Remaining Attributes*

All attributes cannot be described using the SHEF PEDTSEP. However, these attributes are unique to an ESP forecast and need to be provided. The method allows a code to be entered to describe most attributes. Some require a date-time, others an explicit number or string. A description is shown below.

Probability Distributions

CODE	DESCRIPTION
E	Empirical (Default)
N	Normal
L	Log Normal
W	Wakeby
B	Weibull

Weighting Schemes

CODE	DESCRIPTION
EQU	Equal/No Weights (Default)
CPR	CPC Pre Adjustment Technique
CPP	CPC Post Adjustment Technique
ALK	Alaska Technique
YW1	Year Weighting Scheme 1
YW2	Year Weighting Scheme 2
YW3	Year Weighting Scheme 3

Carryover Group

CODE	DESCRIPTION
#	#=alphanumeric string

Date/Time of Extremum

CODE	DESCRIPTION
CCYR-MM-DY HR:MN	(INFORMIX Date Time String)

Date/Time of Carryover

CODE	DESCRIPTION
CCYR-MM-DY HR:MN	(INFORMIX Date Time String)

Data Type of the Time Series

CODE	DESCRIPTION
S	SQME
Q	QINE
N	QIN

Time Step of the Time Series

CODE	DESCRIPTION
24	Daily
1	Hourly
6	6 Hourly

Beginning Year of Historical Data

CODE	DESCRIPTION
#	# = Year

Ending Year of Historical Data

CODE	DESCRIPTION
#	# = Year

Value Used in NDTO(1), NDTO(2), NDIS(1), or NDIS(2)

CODE	DESCRIPTION
#	# = value

Units of the Value Used in NDTO(1), NDTO(2), NDIS(1), or NDIS(2

CODE	DESCRIPTION
C	CFS
F	FEET

(3.3) *Standard PEDTSEP SHEF Constructs*

TYPE PROB	HISTORICAL OBSERVED	HISTORICAL SIMULATED	CONDITIONAL SIMULATED

Flow Volume Over a Variable Period
SUM

QCVHEZZ	QCVPEZZ	QCVFUZZ

Maximum Mean Daily During a Period (Including time of Max)
MXMD

QCVHEUZ	QCVPEUZ	QCVFUUZ

Minimum Mean Daily During a Period (Including time of Min)
MNMD

QCVHEKZ	QCVPEKZ	QCVFUKZ

Maximum Instantaneous Flow During a Period (Including time of Max)
MXIN

QRVHEUZ	QRVPEUZ	QRVFUUZ

Minimum Instantaneous Flow During a Period (Including time of Min)
MNIN

QRVHEKZ	QRVPEKZ	QRVFUKZ

Mean Daily Discharge
MD

QRVHEZZ	QRVPEZZ	QRVFUZZ

Number of Days to Get Above a Value (including value)
NDTO(1)

NDVHEZZ	NDVPEZZ	NDVFUZZ

Number of Days to Get Below a Value (including value)
NDTO(2)

NEVHEZZ	NEVPEZZ	NEVFUZZ

Number of Days Greater than a Value (including value)
NDIS(1)

NFVHEZZ	NFVPEZZ	NFVFUZZ

Number of Days Less Than a Value (including value)
NDIS(2)

NGVHEZZ	NGVPEZZ	NGVFUZZ

Time Series Discharge

6 Hrly	**QRQHEZZ** [Note 3]	**QRQPEZZ**	**QRQFUZZ**
Daily	**QRDHEZZ**	**QRDPEZZ**	**QRDFUZZ**

[Note 3] **The probability in the 7th location of the PEDTSEP needs to be specified for time series information.**

4.0 *CSV Format*

The ESP forecast needs to be converted into its component pieces and placed in a 'CSV' (comma separated value) format. A piece of software has been developed that accepts ESPAPD output and produces the required 'csv' format. The values in the CSV format need to adhere to a metadata format and in the correct position. The software handles this automatically.

There are two metadata formats, one for each category of ESP forecast as described in the section, "ESP Forecast Review". Each value must be separated by a ','.

Category: Single Element in an Analysis Window

```
This format describes various probabilities for a single forecast
element.

Location/Station ID
PEDTSEP
Type of units for analysis window
number of units for analysis window
creation date time
beginning date time or window
carryover date time
carryover group
probability distribution
weighting scheme
historical data type
historical time step
beginning year of historical data
ending year of historical data
ending date time or window
time of extremum
```

```
value for number of days (to/is)
units for value of number of days
quality code
value at probability p_002
value at probability p_004
value at probability p_010
value at probability p_020
value at probability p_040
value at probability p_050
value at probability p_100
value at probability p_200
value at probability p_250
value at probability p_300
value at probability p_400
value at probability p_500
value at probability p_600
value at probability p_700
value at probability p_750
value at probability p_800
value at probability p_900
value at probability p_950
value at probability p_960
value at probability p_980
value at probability p_990
value at probability p_996
value at probability p_998
```

Category: Time Series in an Analysis Window

This format describes a time series of values for a given probability. The probability is described in the PEDTSEP.

```
Location/Station ID
PEDTSEP
creation date time
beginning date time or window
carryover date time
carryover group
probability distribution
weighting scheme
historical data type
historical time step
beginning year of historical data
ending year of historical data
ending date time or window
quality code
value at offset   1
value at offset   2
value at offset   3
value at offset   4
value at offset   5
value at offset   6
value at offset   7
value at offset   8
value at offset   9
value at offset   10
value at offset   11
```

```
value at offset   12
value at offset   13
value at offset   14
value at offset   15
value at offset   16
value at offset   17
value at offset   18
value at offset   19
value at offset   20
value at offset   21
value at offset   22
value at offset   23
value at offset   24
value at offset   25
value at offset   26
value at offset   27
value at offset   28
```

5.0 *Examples*

4.1 ***MXND Maximum Mean Daily & Date/Time of Max***

Attributes	Values for Example
Basin/Segment ID	TEST1BAS
Date/Time of Creation	August 21, 2001 1200
Date/Time of Period	August 28, 2001 1200
Forecast Type	MXND Max Mean Daily
Date/Time of Analysis Window	August 21 - 28 (One Week)
Model Source	Historical Simulated
Extremum	Maximum
Probability Distribution	Normal
Weighting Scheme	Alaska
Date/Time of Carryover	August 06, 2001 1200
Carryover Group ID	ABCDEFG
Data Type of the Time Series	QINE
Time Step of the Time Series	Daily
Beginning Historical Year	1971
Ending Historical Year	1989
Date/Time of Extremum	August 27, 0600 06:00

Value	**None**
Units of Value	**None**
Quality Code	**None**

TESTBAS1,QCVPEUZ,W,1,2001-08-21 12:00, 2001-08-21 12:00, 2001-06-21 12:00,ABCDEFG,N,ALK,QINE,D,1971,1989, 2001-08-28 12:00, 2001-08-27 06:00, , , ,9980,9960,9900,9800,9600,9500,9000,8000,7500,7000,6000,5000,4000,3000,2500,2000,1000,1050, 1040,1020,1010,40,20

---

(5) *Operational Considerations*

The flow of ESP data into the INFORMIX database is described below.

All actions can be set up to execute automatically.

(Step 1)

ESP and/or ESPADP is executed in batch mode (or by the forecaster.) A text output file is created using the probability levels, distribution type, and other user choices. Up to 23 probability levels can be used to describe each forecast type/element in a distribution.

(Step 2)

The text output file from step one is parsed and converted to SHEF/Expanded SHEF. A prototype of this program exists.

NOTE: A better way would be to incorporate a nationally supported option in ESP and ESPADP that would allow the user to output forecasts in SHEF/Enhanced SHEF format.

(Step 3)

The SHEF encoded ESP forecast are copied to a directory or queue. A SHEF/ESP decoder/poster picks it up and parses and posts the data. (A prototype version of the SHEF parser/poster exists.)

---

(6) INFORMIX Table Structure

A sample table schema is shown below. The prime key is marked with an asterick.

```
Single Element

       char     id8[8]                       *
       char     pc[7]                        *
       char     type_units[1]                *
       int      num_units                    *
       datetime year to minute c_dtime       *
       datetime year to minute v1_dtime      *
       datetime year to minute carryo_dtime  *
       char     carryo_group[8]              *
       char     p_distribution[1]            *
       char     weighting[3]                 *
       char     his_data_type[4]             *
       int      his_time_step                *
       int      his_beg_year                 *
       int      his_end_year                 *
       datetime year to minute v2_dtime
       datetime year to minute ext_dtime
       double   nd_value
       char     nd_units[1]
       char     qcode[1]
       double   p_002
       double   p_004
       double   p_010
       double   p_020
       double   p_040
       double   p_050
       double   p_100
       double   p_200
       double   p_250
       double   p_300
       double   p_400
       double   p_500
       double   p_600
       double   p_700
       double   p_750
       double   p_800
       double   p_900
```

```
        double   p_950
        double   p_960
        double   p_980
        double   p_990
        double   p_996
        double   p_998


Time Series

        char     id8[8]                       *
        char     pc[7]                        *
        datetime year to minute c_dtime       *
        datetime year to minute v1_dtime      *
        datetime year to minute carryo_dtime  *
        char     carryo_group[8]              *
        char     p_distribution[1]            *
        char     weighting[3]                 *
        char     his_data_type[4]             *
        int      his_time_step                *
        int      his_beg_year                 *
        int      his_end_year                 *
        datetime year to minute v2_dtime      *
        char     qcode[1]
        double   o_1
        double   o_2
        double   o_3
        double   o_4
        double   o_5
        double   o_6
        double   o_7
        double   o_8
        double   o_9
        double   o_10
        double   o_11
        double   o_12
        double   o_13
        double   o_14
        double   o_15
        double   o_16
        double   o_17
        double   o_18
        double   o_19
        double   o_20
        double   o_21
        double   o_22
        double   o_23
        double   o_24
        double   o_25
        double   o_26
        double   o_27
        double   o_28
```