SAS/QC® 14.1 User's Guide
The MVPMONITOR Procedure

Chapter 13
The MVPMONITOR Procedure

Contents

Overview: MVPMONITOR Procedure

The MVPMONITOR procedure is used in conjunction with the MVPMODEL and MVPDIAGNOSE procedures to monitor multivariate process variation over time, to determine whether the process is stable, and to detect and diagnose changes in a stable process. Collectively, these three procedures are referred to as the *MVP procedures*. See Chapter 10, "Introduction to Multivariate Process Monitoring Procedures," for a description of how the MVP procedures work together, and Chapter 11, "The MVPDIAGNOSE Procedure," and Chapter 12, "The MVPMODEL Procedure," for detailed descriptions of the other MVP procedures.

The MVPMONITOR procedure produces control charts for multivariate process data. It reads data sets that contain statistics and principal component model information such as those created by the MVPMODEL procedure. The MVPMONITOR procedure creates two types of multivariate control chart: T^2 charts and squared prediction error (SPE) charts. It can also produce contribution plots for out-of-control points in the multivariate control charts and univariate control charts of principal component scores.

Multivariate control charts detect unusual variation that would not be uncovered by individually monitoring the process variables with univariate control charts, such as Shewhart charts. A major impetus in the development of multivariate control charts is the inadequacy of individual univariate control charts in handling correlated measurement variables. A multivariate control chart can detect changes in the linear relationships of the variables in addition to their marginal means and variances.

The multivariate control charts that the MVPMONITOR procedure produces are based on principal component models that reduce the dimensionality of the data by projecting the process measurements to a low-dimensional subspace that is defined by a small number of principal components. This subspace is also known as the *model hyperplane*. T^2 charts are used to monitor variation within the model hyperplane, whereas SPE charts show variation from the model hyperplane.

The principal component approach offers several advantages over the construction of the classical T^2 chart:

- It avoids computational issues that arise when the process variables are collinear and their covariance matrix is nearly singular.

- It offers diagnostic tools for interpreting unusual values of T^2.

- By projecting the data to a low-dimensional subspace, it more adequately describes the variation in a multivariate process, which is often driven by a small number of underlying factors that are not directly observable.

Getting Started: MVPMONITOR Procedure

This example illustrates the basic features of the MVPMONITOR procedure by using airline flight delay data available from the U.S. Bureau of Transportation Statistics at `http://www.transtats.bts.gov`. The example applies multivariate process monitoring to flight delays, and it is a continuation of the example in the section "Getting Started: MVPMODEL Procedure" on page 921.

Suppose you want to use a principal component model to create T^2 and SPE charts to monitor the variation in flight delays. These charts are appropriate because the data are multivariate and correlated.

The following statements create a SAS data set named MWflightDelays, which provides daily average delays by airline for flights that originated in the midwestern United States. The data set contains variables for nine airlines: AA (American Airlines), CO (Continental Airlines), DL (Delta Airlines), F9 (Frontier Airlines), FL (AirTran Airways), NW (Northwest Airlines), UA (United Airlines), US (US Airways), and WN (Southwest Airlines).

```
data MWflightDelays;
   format flightDate MMDDYY8.;
   label flightDate='Date';
   input flightDate :MMDDYY8. AA CO DL F9 FL NW UA US WN;
```

```
    datalines;
02/01/07 14.9  7.1  7.9  8.5 14.8  4.5  5.1 13.4  5.1
02/02/07 14.3  9.6 14.1  6.2 12.8  6.0  3.9 15.3 11.4
02/03/07 23.0  6.1  1.7  0.9 11.9 15.2  9.5 18.4  7.6
02/04/07  6.5  6.3  3.9 -0.2  8.4 18.8  6.2  8.8  8.0
02/05/07 12.0 14.1  3.3 -1.3 10.0 13.1 22.8 16.5 11.5
02/06/07 31.9  8.6  4.9  2.0 11.9 21.9 29.0 15.5 15.2
02/07/07 14.2  3.0  2.1 -0.9 -0.6  7.8 19.9  8.6  6.4
02/08/07  6.5  6.8  1.8  7.7  1.3  6.9  6.1  9.2  5.4
02/09/07 12.8  9.4  5.5  9.3 -0.2  4.6  7.6  7.8  7.5
02/10/07  9.4  3.5  1.5 -0.2  2.2  9.9  3.1 12.5  3.0
02/11/07 12.9  5.4  0.9  6.8  2.1  7.9  3.7 10.7  5.6
02/12/07 34.6 15.9  1.8  1.0  4.5 10.2 14.0 19.1  4.9
02/13/07 34.0 16.0  4.4  6.1 18.3  9.1 30.2 46.3 50.6
02/14/07 21.2 45.9 16.6 12.5 35.1 23.8 40.4 43.6 35.2
02/15/07 46.6 36.3 23.9 20.8 30.4 24.3 30.3 59.9 25.6
02/16/07 31.2 20.8 15.2 20.1  9.1 12.9 22.9 36.4 16.4
;
```

The observations for a given date are the average flight delays in minutes for flights that departed from the midwestern United States. For example, on February 2, 2007, F9 (Frontier Airlines) flights departed 6.2 minutes late on average.

Creating a Multivariate Control Chart in a Phase I Situation

In a Phase I analysis you first perform a principal component analysis (PCA) of the data. Then you can use control charts to determine whether the data that you use to build the principal component model indicate a stable multivariate process. The MVPMONITOR procedure creates multivariate control charts from T^2 and SPE statistics computed from a principal component model that the MVPMODEL procedure produced. This example uses the model built in the section "Building a Principal Component Model" on page 924 in Chapter 12, "The MVPMODEL Procedure."

The following statements fit the model:

```
proc mvpmodel data=MWflightDelays ncomp=3 noprint
              out=mvpair outloadings=mvpairloadings;
   var AA CO DL F9 FL NW UA US WN;
run;
```

The NCOMP= option requests a principal component model that contains three principal components. The OUT= option creates a data set that contains the original data, the principal component scores, and the T^2 and SPE statistics. The OUTLOADINGS= data set contains the variances and loadings for the principal components.

The following statements produce the multivariate control charts:

```
ods graphics on;
proc mvpmonitor history=mvpair loadings=mvpairloadings;
   time flightDate;
   tsquarechart / contributions;
   spechart / contributions;
run;
```

The HISTORY= option specifies the input data set. The LOADINGS= option specifies the data set that contains the principal component model information. The TSQUARECHART statement requests a T^2 chart,

and the SPECHART statement requests an SPE chart. The CONTRIBUTIONS options that are specified in the TSQUARECHART and SPECHART statements request contribution plots for all out-of-control points in the charts. The TIME statement specifies that the variable flightDate provide the chronological ordering of the observations.

Figure 13.1 shows the T^2 chart.

Figure 13.1 Multivariate Control Chart for T^2 Statistics
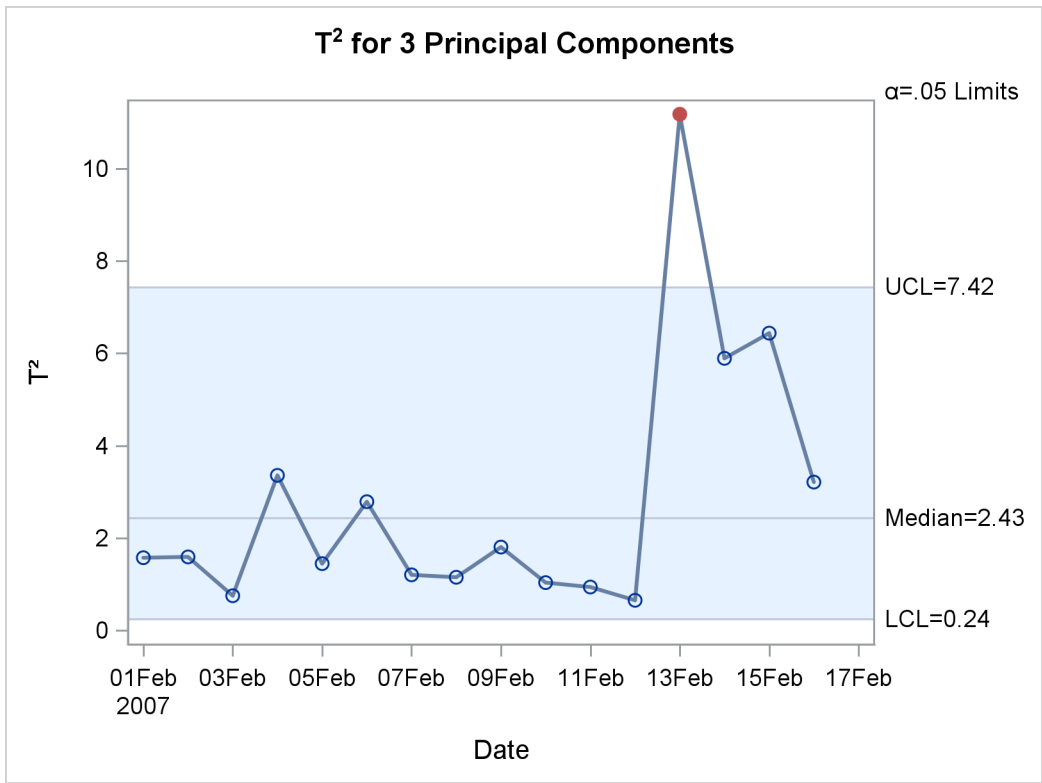
The T^2 chart shows an out-of-control point on February 13, 2007. On this day, a strong winter storm battered the midwestern United States. To see which variables contributed to this statistic, you can use the contribution plot shown in Figure 13.2.

Figure 13.2 Contribution Plot
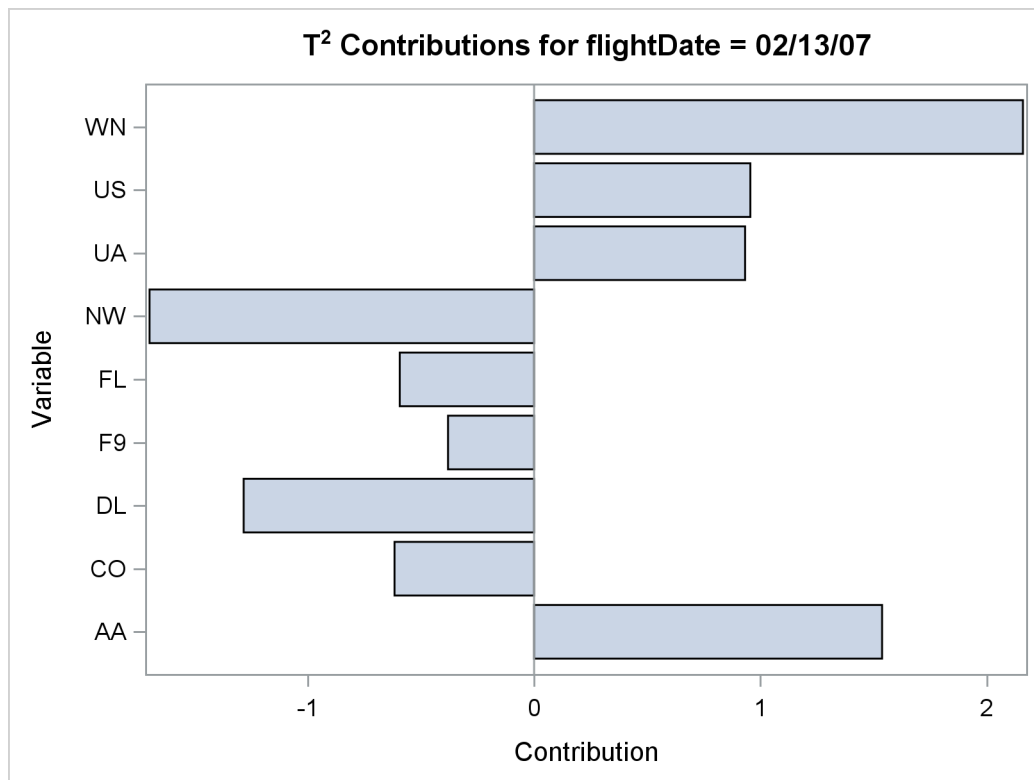
The contribution plot shows that the variables WN, AA, NW, and DL are the major contributors to the out-of-control point.

Figure 13.3 shows the SPE chart.

Figure 13.3 Multivariate Control Chart for SPE Statistics
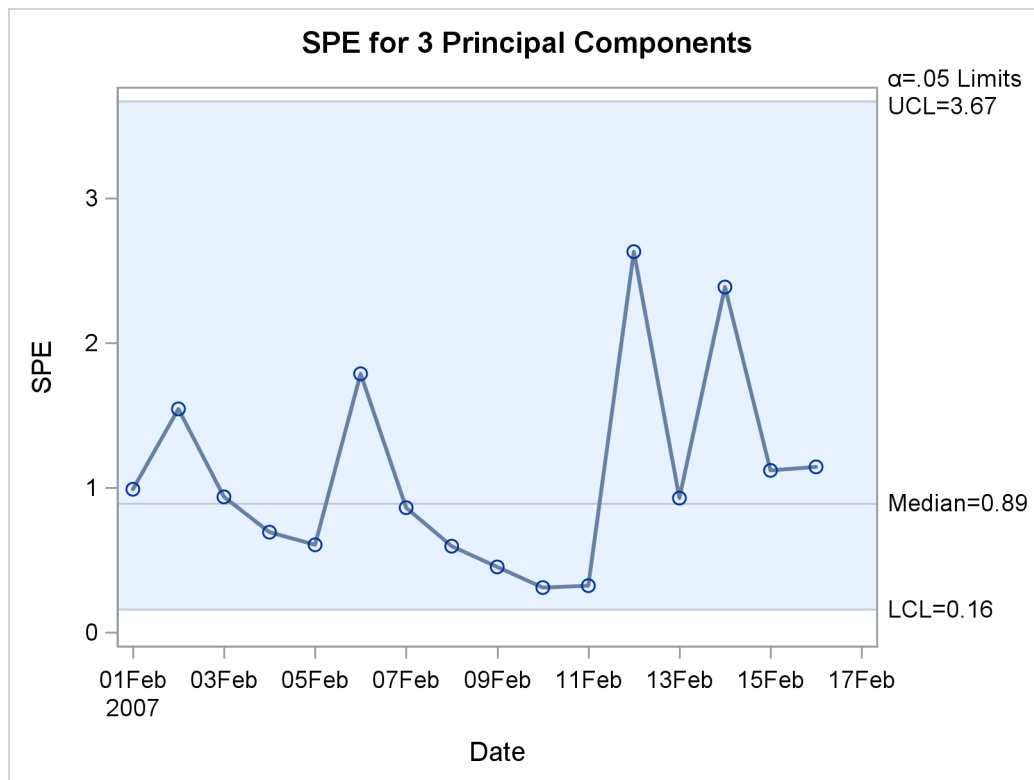
There are no out-of-control points in the SPE chart. This indicates that the unusual point displayed in the T^2 chart represents a departure from the variation described by the principal component model that lies within the model hyperplane.

Syntax: MVPMONITOR Procedure

PROC MVPMONITOR < *options* > ;
    BY *variables* ;
    ID *variable* ;
    SCORECHART < */ options* > ;
    SERIES *variable* ;
    SPECHART < */ options* > ;
    TIME *variable* ;
    TSQUARECHART < */ options* > ;

The following sections describe the PROC MVPMONITOR statement and then describe the other statements in alphabetical order.

PROC MVPMONITOR Statement

> **PROC MVPMONITOR** < *options* > ;

The PROC MVPMONITOR statement invokes the MVPMONITOR procedure and specifies input and output data sets. You can specify the following *options*:

DATA=_SAS-data-set_

specifies an input SAS data set that contains process measurement data for a Phase II analysis. If you specify a DATA= data set, you must also specify a LOADINGS= data set. You cannot specify the HISTORY= or TABLE= option together with the DATA= option. See the section "DATA= Data Set" on page 969 for details about DATA= data sets.

HISTORY=_SAS-data-set_

specifies an input SAS data set that contains process variable data augmented with principal component scores, multivariate summary statistics, and other calculated values. This data set is used for a Phase I analysis. Usually, a HISTORY= data set is created as an OUT= data set from the MVPMODEL procedure. You cannot specify the DATA= or TABLE= option with the HISTORY= option. See the section "HISTORY= Data Set" on page 969 for details about HISTORY= data sets.

LOADINGS=_SAS-data-set_

specifies an input SAS data set that contains eigenvalues, principal component loadings, and process variable means and standard deviations that are used to compute principal component scores and multivariate summary statistics for a Phase II analysis. Usually, the LOADINGS= data set is produced by the MVPMODEL procedure as an OUTLOADINGS= data set. See the section "LOADINGS= Data Set" on page 970 for details about LOADINGS= data sets.

MISSING=AVG | NONE

specifies how to handle observations that have missing process variable values in the DATA= data set. The option MISSING=AVG specifies that missing values for a given variable be replaced by the average of the nonmissing values for that variable. The default is MISSING=NONE, which excludes observations that have missing values for any of the process variables from the analysis.

OUTHISTORY=_SAS-data-set_

OUT=_SAS-data-set_

creates an output data set that contains all the original data from the input data set, principal component scores, and multivariate summary statistics. See the section "OUTHISTORY= Data Set" on page 973 for details. You can produce an OUTHISTORY= data set only when you specify a DATA= input data set.

PREFIX=_name_

specifies the prefix to use to identify variables that contain principal component scores in the HISTORY= data set. For example, if you specify PREFIX=ABC, PROC MVPMONITOR attempts to read the score variables ABC1, ABC2, ABC3, and so on. The default PREFIX= value is Prin, which is the default score variable prefix the that PROC MVPMODEL uses when it creates an OUT= data set. If you use an OUT= data set from MVPMODEL as a HISTORY= data set, the PREFIX= value must match the PROC MVPMODEL PREFIX= value that is specified when the OUT= data set is created.

NOTE: The number of characters in the prefix plus the number of digits that are required to enumerate the principal components must not exceed the maximum name length defined by the VALIDVAR-NAME= system option.

RPREFIX=_name_

specifies the prefix to use to identify variables that contain residuals in the HISTORY= data set. Residual variable names are formed by appending process variable names to the prefix. The default RPREFIX= value is R_, which is the default residual variable prefix that PROC MVPMODEL uses when it creates an OUT= data set. If you use an OUT= data set from PROC MVPMODEL as a HISTORY= data set, the RPREFIX= value must match the PROC MVPMODEL RPREFIX= value that is specified when the OUT= data set is created.

If the combined length of the residual prefix and a process variable name exceeds the maximum name length defined by the VALIDVARNAME= system option, characters are removed from the middle of the process variable name before it is appended to the residual prefix. For example, if you specify RPREFIX=_Residual__ (nine characters), the maximum variable name length is 32, and there is a process variable named PrimaryThermometerReading (25 characters), then two characters are dropped from the middle of the process variable name. The resulting residual variable name is Residual_PrimaryThermeterReading.

TABLE=_SAS-data-set_

specifies an input SAS data set that contains summary information from a score chart, SPE chart, or T^2 chart. You can produce a TABLE= data set by specifying the OUTTABLE= option in a SCORECHART, SPECHART, or TSQUARECHART statement. You can use a TABLE= input data set to display a previously computed control chart. You cannot specify the DATA= or HISTORY= option together with the TABLE= option. See the section "TABLE= Data Set" on page 971 for details.

BY Statement

BY _variables_ **;**

You can specify a BY statement with PROC MVPMONITOR to obtain separate analyses of observations in groups that are defined by the BY variables. When a BY statement appears, the procedure expects the input data set to be sorted in order of the BY variables. If you specify more than one BY statement, only the last one specified is used.

If your input data set is not sorted in ascending order, use one of the following alternatives:

- Sort the data by using the SORT procedure with a similar BY statement.

- Specify the NOTSORTED or DESCENDING option in the BY statement for the MVPMONITOR procedure. The NOTSORTED option does not mean that the data are unsorted but rather that the data are arranged in groups (according to values of the BY variables) and that these groups are not necessarily in alphabetical or increasing numeric order.

- Create an index on the BY variables by using the DATASETS procedure (in Base SAS software).

For more information about BY-group processing, see the discussion in *SAS Language Reference: Concepts*. For more information about the DATASETS procedure, see the discussion in the *Base SAS Procedures Guide*.

ID Statement

ID *variables* ;

The values of the ID *variables* are displayed in tooltips associated with points on a T^2 or SPE chart when you create HTML output and specify the IMAGEMAP option in the ODS GRAPHICS statement. See Chapter 21, "Statistical Graphics Using ODS" (*SAS/STAT User's Guide*), for details.

SERIES Statement

SERIES *variable* ;

The SERIES *variable* uniquely identifies a series of observations in the input data set to be plotted in a control chart. You must specify a SERIES statement when the input data set contains more than one observation that has the same TIME variable value. See Example 13.1 for an illustration of how to use the SERIES statement.

SCORECHART Statement

SCORECHART < / *options* > ;

The SCORECHART statement produces univariate control charts of principal component scores. In a Phase I analysis, the scores are computed by PROC MVPMODEL and read from a HISTORY= data set. In a Phase II analysis, PROC MVPMONITOR computes the scores from process data contained in a DATA= data set and information about the principal component model that is specified in the LOADINGS= data set.

Table 13.1 summarizes the *options* available in the SCORECHART statement.

Table 13.1 SCORECHART Statement Options

Option	Description
COMP=	Specifies the principal components whose scores are plotted
EXCHART	Specifies that a score chart be displayed only if there are points lying outside the control limits
INTERVAL=	Specifies the time interval between consecutive positions on a score chart
NOCHART	Suppresses creation of score charts
NOHLABEL	Suppresses the horizontal axis label of a score chart
NOVLABEL	Suppresses the vertical axis label of a score chart
NPANELPOS=	Specifies the number of horizontal axis plotting positions per panel
ODSFOOTNOTE=	Adds a footnote to a score chart
ODSFOOTNOTE2=	Adds a secondary footnote to a score chart
ODSTITLE=	Specifies a title for a score chart
ODSTITLE2=	Specifies a secondary title for a score chart
OUTTABLE=	Creates a SAS data set that contains a summary of a score chart
OVERLAY=	Overlays separate series of observations or scores for multiple principal components in a single score chart

Table 13.1 *(continued)*

Option	Description
SERIESVALUE=	Specifies SERIES variable values that select sequences to be plotted
SIGMAS=	Specifies the multiples of a score's standard deviation used to compute score chart limits
TOTPANELS=	Specifies the number of panels used to display a score chart

The following options are available only in the SCORECHART statement. For detailed descriptions of options common to the SCORECHART, SPECHART, and TSQUARECHART statements, see the section "Common Chart Statement Options" on page 963.

COMP=*value-list* **| ALL**

specifies the principal components whose scores are plotted. You can specify the following values:

value-list plots the principal components whose numbers (1, 2, ...) appear in *value-list*.

ALL plots all the principal components whose scores are in the input data set.

If you do not specify this option, scores for only the first principal component are plotted.

OVERLAY=COMP | SERIES

requests that multiple sequences of principal component scores be overlaid on a single control chart. You can specify the following values:

COMP overlays the scores for different principal components on the chart. The COMP= option determines which principal components are included.

SERIES overlays sequences of scores defined by the SERIES variable and selected by the SERIESVALUE= option on the chart. This value is applicable only when a TIME variable is specified and the input data set contains multiple scores for each principal component that have the same time value.

The value you specify in this option determines the number of separate control charts that are produced. For example, if you select p principal components and n SERIESVALUE= values, PROC MVPMONITOR produces

- n charts if you specify OVERLAY=COMP,
- p charts if you specify OVERLAY=SERIES, and
- $p \times n$ charts if you do not specify the OVERLAY= option.

SIGMAS=*k*

specifies the width of the control limits in terms of the multiple k of a score's standard deviation. By default, $k=3$.

SPECHART Statement

SPECHART < / *options* > ;

The SPECHART statement produces a control chart of the squared prediction error (SPE) statistics based on a principal component model that the MVPMODEL procedure computes. In a Phase I analysis, the SPE statistics are computed by PROC MVPMODEL and read from the HISTORY= data set. In a Phase II analysis, the SPE statistics are computed by the MVPMONITOR procedure from information about the principal component model that is specified in the LOADINGS= data set.

Table 13.2 summarizes the *options* available in the SPECHART statement. For detailed descriptions, see the section "Common Chart Statement Options" on page 963.

Table 13.2 SPECHART Statement Options

Option	Description
ALPHA=	Specifies the probability used to compute SPE chart limits
CONTRIBUTIONS	Creates contribution plots for points outside control limits
EXCHART	Specifies that an SPE chart be displayed only if there are points lying outside control limits
INTERVAL=	Specifies the time interval between consecutive positions on an SPE chart
NOCHART	Suppresses creation of an SPE chart
NOHLABEL	Suppresses the horizontal axis label of an SPE chart
NOVLABEL	Suppresses the vertical axis label of an SPE chart
NPANELPOS=	Specifies the number of horizontal axis plotting positions per panel
ODSFOOTNOTE=	Adds a footnote to an SPE chart
ODSFOOTNOTE2=	Adds a secondary footnote to an SPE chart
ODSTITLE=	Specifies a title for an SPE chart
ODSTITLE2=	Specifies a secondary title for an SPE chart
OUTTABLE=	Creates a SAS data set that contains a summary of an SPE chart
OVERLAY	Specifies that separate sequences of observations be plotted in a single SPE chart
SERIESVALUE=	Specifies SERIES variable values selecting sequences to be plotted
TOTPANELS=	Specifies the number of panels used to display an SPE chart

TIME Statement

TIME *variable* ;

The TIME *variable* is a numeric variable that provides the chronological order or time values for measurements in a DATA=, HISTORY=, or TABLE= input data set. The values of the TIME *variable* are displayed on the horizontal axis of control charts. If no TIME statement is specified, the observation numbers from the input data set are displayed on the horizontal axis.

When the input data set contains more than one observation with the same TIME value, you must identify the sequences of points to be plotted on the control charts by specifying a SERIES variable.

TSQUARECHART Statement

TSQUARECHART < / *options* > ;

The TSQUARECHART statement produces a T^2 chart based on a principal component model that the MVPMODEL procedure computes. In a Phase I analysis, the T^2 statistics are computed by PROC MVP-MODEL and read from the HISTORY= data set. In a Phase II analysis, the T^2 statistics are computed by the MVPMONITOR procedure from information about the principal component model specified in the LOADINGS= data set.

Table 13.3 summarizes the *options* available in the TSQUARECHART statement. For detailed descriptions, see the section "Common Chart Statement Options" on page 963.

Table 13.3 TSQUARECHART Statement Options

Option	Description
ALPHA=	Specifies the probability used to compute T^2 chart limits
CONTRIBUTIONS	Creates contribution plots for points outside the control limits
EXCHART	Specifies that a T^2 chart be displayed only if there are points lying outside the control limits
INTERVAL=	Specifies the time interval between consecutive positions on a T^2 chart
LIMITDIST=	Specifies the distribution that is used to compute the control limits
NOCHART	Suppresses creation of a T^2 chart
NOHLABEL	Suppresses the horizontal axis label of a T^2 chart
NOVLABEL	Suppresses the vertical axis label of a T^2 chart
NPANELPOS=	Specifies the number of horizontal axis plotting positions per panel
ODSFOOTNOTE=	Adds a footnote to a T^2 chart
ODSFOOTNOTE2=	Adds a secondary footnote to a T^2 chart
ODSTITLE=	Specifies a title for a T^2 chart
ODSTITLE2=	Specifies a secondary title for a T^2 chart
OUTTABLE=	Creates a SAS data set that contains a summary of a T^2 chart
OVERLAY	Specifies that separate sequences of observations be plotted on a single T^2 chart
SERIESVALUE=	Specifies SERIES variable values selecting sequences to be plotted
TOTPANELS=	Specifies the number of panels used to display a T^2 chart

LIMITDIST=BETA | CHISQ | F

specifies the distribution that is used to compute control limits for the T^2 chart. You can specify the following values:

BETA   selects a beta distribution

CHISQ   selects a χ^2 distribution

F   selects an *F* distribution

See the section "Computing T^2 Control Limits" on page 966 for a description of how T^2 chart control limits are computed.

Common Chart Statement Options

You can specify the following *options* after a slash (/) in a SCORECHART, SPECHART, or TSQUARECHART statement (unless noted otherwise).

ALPHA=*value*
> specifies the probability used to compute control limits for an SPE or T^2 chart. If you specify ALPHA=α, the control limits are computed so that the probability is α that the statistic exceeds its control limits. The *value* of α can range from 0 to 1. By default, α is 0.05. This option is not available in the SCORECHART statement.

CONTRIBUTIONS < **(***contribution-options***)** >
> creates a contribution plot for each point that falls outside the control limits of the chart. You can specify the following *contribution-options* in parentheses:

> **MAXNPLOTS=***n*
> > specifies the maximum number, *n*, of contribution plots to produce. When *n* is less than the number of points outside the control limits, contribution plots are produced for the first *n* out-of-control points.

> **MAXNVARS=***n*
> > specifies that only the *n* contributions with the greatest magnitudes be displayed in the contribution plots. For each out-of-control point, the *n* variables that contribute the most to that value of the statistic are displayed. By default, all variable contributions are displayed.

> > This option is not available in the SCORECHART statement.

EXCHART
> specifies that a control chart be displayed only when one or more points lie outside the control limits.

INTERVAL=DAY | DTDAY | HOUR | MINUTE | MONTH | QTR | SECOND
> specifies the natural time interval between consecutive TIME variable positions when a time, date, or datetime format is associated with the TIME variable. By default, the INTERVAL= option uses the number of positions per panel that you specify with the NPANELPOS= option. The default time interval keywords for various time formats are shown in Table 13.4.

Table 13.4 Default Time Intervals

Time Format	Default Interval
DATE	DAY
DATETIME	DTDAY
DDMMYY	DAY
HHMM	HOUR
HOUR	HOUR
MMDDYY	DAY
MMSS	MINUTE
MONYY	MONTH
TIME	SECOND
TOD	SECOND
WEEKDATE	DAY
WORDDATE	DAY
YYMMDD	DAY
YYQ	QTR

You can use the INTERVAL= option to modify the effect of the NPANELPOS= option, which specifies the number of positions per panel. The INTERVAL= option enables you to match the scale of the horizontal axis to the scale of the TIME variable without having to associate a different format with the TIME variable.

For example, suppose your formatted time values span 100 days and a datetime format is associated with the TIME variable. Since the default interval for the datetime format is dtday and because NPANELPOS=50 by default, the chart is displayed with two panels.

Now suppose your data span 100 hours and a datetime format is associated with the TIME variable. The chart for these data is created in a single panel, but the data occupy only a small fraction of the chart because the scale of the data (hours) does not match that of the horizontal axis (days). If you specify INTERVAL=HOUR, the horizontal axis is scaled for 50 hours, matching the scale of the data, and the chart is displayed with two panels.

NOCHART

suppresses display of the control chart. You can use the NOCHART option with the CONTRIBUTIONS option to produce contribution plots for out-of-control points without displaying the control chart. You can use the NOCHART option with the OUTTABLE= option to save a summary of the control chart to an output data set without displaying the chart.

NOHLABEL

suppresses the horizontal axis label in the control chart.

NOVLABEL

suppresses the vertical axis label in the control chart.

NPANELPOS=n

NPANEL=n

specifies the number of horizontal axis plotting positions per panel in the chart. You typically specify the NPANELPOS= option to display more points in a panel than the default number, which is 50.

You can specify a positive or negative value for *n*. The absolute value of *n* must be at least 5. If *n* is positive, the number of positions is adjusted so that it is approximately equal to *n* and so that all panels display approximately the same number of positions. If *n* is negative, then no balancing is done, and each panel (except possibly the last) displays approximately |*n*| positions.

ODSFOOTNOTE=FOOTNOTE | FOOTNOTE1 | '*string***'**

adds a footnote to the chart. If you specify the FOOTNOTE (or FOOTNOTE1) keyword, the value of the SAS FOOTNOTE statement is used as the chart footnote. If you specify a quoted string, that string is used as the footnote. The quoted string can contain the following escaped characters, which are replaced with the values indicated:

\n   is replaced by the TIME variable name.

\l   is replaced by the TIME variable label (or name if the TIME variable has no label).

ODSFOOTNOTE2=FOOTNOTE2 | '*string***'**

adds a secondary footnote to the chart. If you specify the FOOTNOTE2 keyword, the value of the SAS FOOTNOTE2 statement is used as the secondary chart footnote. If you specify a quoted string, that string is used as the secondary footnote. The quoted string can contain the following escaped characters, which are replaced with the values indicated:

\n   is replaced by the TIME variable name.

\l   is replaced by the TIME variable label (or name if the TIME variable has no label).

ODSTITLE=TITLE | TITLE1 | NONE | DEFAULT | '*string***'**

specifies a title for the chart. You can specify the following values:

TITLE (or TITLE1)   uses the value of the SAS TITLE statement as the chart title.

NONE   suppresses all titles from the chart.

DEFAULT   uses the default title.

If you specify a quoted string, that string is used as the graph title. The quoted string can contain the following escaped characters, which are replaced with the values indicated:

\n   is replaced by the TIME variable name.

\l   is replaced by the TIME variable label (or name if the analysis variable has no label).

ODSTITLE2=TITLE2 | '*string***'**

specifies a secondary title for the chart. If you specify the TITLE2 keyword, the value of the SAS TITLE2 statement is used as the secondary chart title. If you specify a quoted string, that string is used as the secondary title. The quoted string can contain the following escaped characters, which are replaced with the values indicated:

\n   is replaced by the TIME variable name.

\l   is replaced by the TIME variable label (or name if the analysis variable has no label).

OUTTABLE=SAS-data-set

> creates an output SAS data set that contains the information plotted in the control chart, including the statistic values and control limits. See the section "OUTTABLE= Data Set" on page 973 for a description of the OUTTABLE= data set.

OVERLAY

> specifies that the separate sequences of observations defined by the SERIES variable and selected by the SERIESVALUE= option be plotted in a single control chart. By default, each sequence is plotted in a separate chart. The OVERLAY option is applicable only when a TIME variable is specified and the input data set contains multiple observations that have the same time value.

> This option is not available in the SCORECHART statement. You can use the OVERLAY= option in the SCORECHART statement to overlay control charts of principal component scores.

SERIESVALUE=value-list

> specifies a list of values of the SERIES variable that define one or more sequences of observations to be plotted. If the SERIES variable is a character variable, the *value-list* must be a list of quoted strings. By default, a series is plotted for each unique value of the SERIES variable. The SERIESVALUE= option is applicable only when a TIME variable is specified and the input data set contains multiple observations that have the same time value.

TOTPANELS=n

> specifies the number of panels to use to display the chart. By default, the number of panels is determined by the value that you specify in the NPANELPOS= option. If you specify both the TOTPANELS= and NPANELPOS= options, the TOTPANELS= value takes precedence.

Details: MVPMONITOR Procedure

Computing T^2 Control Limits

The control limits for the T^2 chart are the same for all the T^2 statistics on the chart. The control limits are computed based on one of the following distributions:

- a beta distribution

$$T_i^2 \sim \frac{(n-1)^2}{n} B\left(\frac{j}{2}, \frac{n-j-1}{2}\right) \qquad j \geq 2, n \geq j+1$$

- a χ^2 distribution

$$T_i^2 \sim \chi^2(j) \qquad j \geq 2, n \geq j+1$$

- an F distribution

$$T_i^2 \sim \frac{j(n+1)(n-1)}{n(n-j)} F(j, n-j) \qquad j \geq 2, n \geq j+1$$

where i is the observation, j is the number of principal components in the model, and n is the number of observations used to build the principal component model.

The upper control limit is computed as the $(1 - \frac{\alpha}{2})$ quantile of the distribution, and the lower control limit is computed as the $\frac{\alpha}{2}$ quantile. You can specify the ALPHA= option in the TSQUARECHART statement to specify α.

You can specify the LIMITDIST= option in the TSQUARECHART statement to select the distribution that is used to compute the control limits. A beta distribution is used by default. See Tracy, Young, and Mason (1992) for a discussion of the conditions under which each distribution is applicable.

See the section "Computing the T^2 and SPE Statistics" on page 938 for details of computing the T^2 statistic based on a principal component model.

Computing SPE Control Limits

The SPE chart plots the sum of squares of the residuals from the principal component model. If either $j = p$ or the data matrix has rank less than p, then the SPE statistic is not defined and an SPE chart is not produced. The SPE statistic for observation i is denoted as

$$Q_i = \sum_{k=1}^{p} e_{ik}^2$$

where p is the number of variables and e_{ik} is the ith observation for the kth variable in the error matrix, E, in the principal component model

$$\mathbf{X} = \mathbf{TP'} + \mathbf{E}$$

The distribution of Q_i has been approximated in the literature under different conditions. Two methods of computing control limits for Q_i are implemented by the MVPMONITOR procedure. One method is used when the data that are used to build the principal component model consist of a single measurement per time point. The other method is used when there are multiple measurements per time point (Jensen and Solomon 1972; Nomikos and MacGregor 1995).

One Observation per Time Point

When there is a single observation at each time point, the data matrix $\mathbf{X}$ is $n \times p$, with exactly one observation at each time point in the input data set. The derivation of the control limits uses the central limit theorem approach of Jensen and Solomon (1972). They begin by defining $\theta_i = \sum_{k=j+1}^{p} \lambda_k^i$, $i = 1, 2, 3$, where λ_k is the kth eigenvalue from the principal component model.

Then the quantity

$$z = \frac{\theta_1 \left[\left(\frac{Q}{\theta_1} \right)^{h_0} - 1 - \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]}{\sqrt{2\theta_2 h_0^2}}$$

is distributed $N(0, 1)$ as $n \to \infty$, where h_0 is defined as $1 - \frac{2\theta_1\theta_3}{3\theta_2^2}$. The upper control limit for all Q_i is then computed by

$$Q_{i,1-\frac{\alpha}{2}} = \theta_1 \left[1 + \frac{z_{(1-\alpha/2)}\sqrt{2\theta_2 h_0^2}}{\theta_1} + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}}$$

where $z_{(1-\alpha/2)}$ is the $(1 - \frac{\alpha}{2})$ percentile of the standard normal distribution. The lower control limit is obtained similarly by using $\frac{\alpha}{2}$. You can specify α by using the ALPHA= option in the SPECHART statement.

Multiple Observations per Time Point

When there are multiple observations at a time value in an input data set, a different approximation of the SPE distribution is used to compute control limits. The approximate distribution at time i is the scaled chi-square distribution,

$$\frac{s_i^2}{2\bar{x}_i} \chi^2_{\frac{2\bar{x}_i^2}{s_i^2}}$$

where $\bar{x}_i$ and s_i^2 are the mean and variance, respectively, of the SPE statistics at time i. The upper control limit for all observations at time point i is computed as the $(1 - \frac{\alpha}{2})$ percentile of the scaled chi-square distribution:

$$SPE_{i,1-\frac{\alpha}{2}} = \frac{s_i^2}{2\bar{x}_i} \chi^2_{\frac{2\bar{x}_i^2}{s_i^2},1-\frac{\alpha}{2}}$$

Similarly the lower control limit is computed from the $\frac{\alpha}{2}$ percentile. You can specify α by using the ALPHA= option in the SPECHART statement.

For more information about the distribution approximations, see Nomikos and MacGregor (1995) and Jackson and Mudholkar (1979).

Contribution Plots

One way to diagnose the behavior of out-of-control points in multivariate charts is to use contribution plots (Miller, Swanson, and Heckler 1998). These plots tell you which variables contribute to the distance between the points in an SPE or T^2 chart and the sample mean of the data.

A contribution plot is a bar chart of the contributions of the process variables to the statistic. For the ith SPE statistic, the contribution of the kth variable is the kth entry of the vector $\mathbf{e}_i$, which is computed as

$$\mathbf{e}_i = \mathbf{x}_i \left(\mathbf{I} - \mathbf{P}_j \mathbf{P}'_j \right)$$

where $\mathbf{e}_i$ is the vector of errors from the principal component model for observation i and $\mathbf{x}_i$ is the ith observation. The contributions to the ith T^2 statistic are computed in the same way as the entries of the vector

$$\mathbf{T}_i^2 = \mathbf{x}_i \mathbf{P}_j \mathbf{L}^{-1} \mathbf{P}'_j$$

where $\mathbf{P}_j$ is the matrix of the first j eigenvectors and $\mathbf{L}$ is the diagonal matrix of the first j eigenvalues.

Input Data Sets

The MVPMONITOR procedure accepts a single primary input data set of one of three types.

- A DATA= data set contains new process data to be analyzed by using an existing PCA model (Phase II analysis).

- A HISTORY= data set contains process data and the accompanying scores, residuals, and statistics produced by applying a PCA model. The process data can be the original data that was used to create the model (Phase I analysis) or subsequent data that was analyzed by using a previously created model (Phase II analysis).

- A TABLE= data set contains a summary of score charts, SPE charts, or T^2 charts, which consists of the statistics, control limits, and other information.

These options are mutually exclusive. If you do not specify an option identifying a primary input data set, PROC MVPMONITOR uses the most recently created SAS data set as a DATA= data set.

When you specify a DATA= data set, you must also specify a LOADINGS= data set that contains loadings and other information describing the PCA model. When you specify a HISTORY= data set, you must also specify a LOADINGS= data set if you specify the CONTRIBUTIONS option in a TSQUARECHART statement.

DATA= Data Set

A DATA= data set provides the process measurement data for a Phase II analysis. In addition to the process variables, a DATA= data set can include the following:

- BY variables

- ID variables

- a SERIES variable

- a TIME variable

When you specify a DATA= data set, you must also specify a LOADINGS= data set that contains the loadings for the principal component model that describes the variation of the process. These loadings are used to score the new data from the DATA= data set. The process variables in the LOADINGS= data set must have the same names as those in the DATA= data set.

HISTORY= Data Set

A HISTORY= data set provides the input data set for a Phase I or Phase II analysis. In addition to the original process variables, it contains principal component scores, residuals, SPE and T^2 statistics, and a count of the observations that are used to construct the principal component model, as summarized in Table 13.5.

Table 13.5 Variables in the HISTORY= Data Set

Variable	Description
Prin1–Prin*j*	Principal component scores
R_*var*1–R_*varp*	Residuals
NOBS	Number of observations used to build the principal component model
SPE	Squared prediction error (SPE)
TSQUARE	T^2 statistic computed from principal component scores

A HISTORY= data set must include variables that contain principal component scores. The score variables names must consist of a common prefix followed by the numbers 1, 2, ..., *j*, where *j* is the number of principal components. By default, the common prefix is Prin. You can use the PREFIX= option to specify another prefix for score variables.

If the number of principal components is less than the total number of process variables, the HISTORY= data set should also contain residual variables. A residual variable name consists of a common prefix followed by the corresponding process variable name. The default residual variable prefix is R_. For example, if the process variables are A, B, and C, the default residual variable names are R_A, R_B, and R_C. You can use the RPREFIX= option to specify a different residual variable prefix.

NOTE: Usually you create a HISTORY= data set by specifying the PROC MVPMODEL OUT= option or the PROC MVPMONITOR OUTHISTORY= option. If the PREFIX= or RPREFIX= option is used when such an output data set is created, you must specify the same prefixes to identify the score and residual variables when you read it as a HISTORY= data set.

LOADINGS= Data Set

The LOADINGS= data set contains the following information about the principal component model:

- eigenvalues of the correlation or covariance matrix used to construct the model

- principal component loadings

- process variable means used to center the variable values

- process variable standard deviations used to scale the variable values

You can produce a LOADINGS= data set by using the PROC MVPMODEL OUTLOADINGS= option. Table 13.6 lists the variables that are required in a LOADINGS= data set.

Table 13.6 Variables in the LOADINGS= Data Set

Variable	Description
VALUE	The value contained in *process variables* for a given observation
NOBS	Number of observations used to build the principal component model
_PC_	Principal component number; 0 for the observation that contains eigenvalues
process variables	Values associated with the process variables

Valid values for the _VALUE_ variable are as follows:

EIGEN   eigenvalues from the principal component analysis

LOADING   principal component loadings

MEAN   process variable means

STD   process variable standard deviations

The LOADINGS= data set contains one EIGEN observation and j LOADING observations, where j is the number of principal components in the model. The presence of a MEAN observation indicates that the process variables were centered when the principal component model was constructed, and the presence of a STD observation indicates that the process variables were scaled when the principal component model was constructed. The means and standard deviations are used to center and scale new data in a Phase II analysis.

TABLE= Data Set

A TABLE= data set contains a summary of one or more score charts, SPE charts, or T^2 control charts. Usually, you create a TABLE= data set by specifying the OUTTABLE= option in a SCORECHART, SPECHART, or TSQUARECHART statement. Each type of TABLE= data set contains different variables, and when you specify a TABLE= data set you can only specify chart statements of the corresponding type. For example, if you use a TABLE= data set that contains SPE chart summary data, you cannot specify a SCORECHART or TSQUARECHART statement.

You can use a TABLE= data set to display previously created control charts or to specify custom control limits by computing your own _LCL_ and _UCL_ values.

Table 13.7, Table 13.8, and Table 13.9 list the variables that are contained in the three types of TABLE= data set.

NOTE:

1. SPE chart and T^2 chart TABLE= data sets contain one observation per time value. Score chart TABLE= data sets contain one observation for each principal component per time value.

2. SPE chart and T^2 chart TABLE= data sets contain residual variables corresponding to the process variables. Each residual variable has the same name as the corresponding process variable

Table 13.7   Score Chart TABLE= Data Set Variables

Variable	Description
COMP	Principal component number
EXLIM	Flag that indicates control limit was exceeded
LCL	Lower control limit
MEAN	Center line
SCORE	Principal component score
series	Optional SERIES variable
SIGMAS	Multiple of score standard deviation used to compute control limits
time	Optional TIME variable
UCL	Upper control limit

Table 13.8 SPE Chart TABLE= Data Set Variables

Variable	Description
ALPHA	Probability (α) of exceeding control limits
EXLIM	Flag to indicate control limit was exceeded
LCL	Lower control limit
MEDIAN	Center line
residuals	Residual variables
series	Optional SERIES variable
SPE	Squared prediction error (SPE) statistic
time	Optional TIME variable
UCL	Upper control limit

Table 13.9 T^2 Chart TABLE= Data Set Variables

Variable	Description
ALPHA	Probability (α) of exceeding control limits
EXLIM	Flag to indicate control limit was exceeded
LCL	Lower control limit
MEDIAN	Center line
residuals	Residual variables
series	Optional SERIES variable
time	Optional TIME variable
TSQUARE	T^2 statistic (TSQUARECHART statement only)
UCL	Upper control limit

Output Data Sets

OUTHISTORY= Data Set

The OUTHISTORY= data set contains all the variables in the input data set plus new variables that contain the principal component scores, residuals, and other computed values listed in Table 13.10.

Table 13.10 Computed Variables in the OUTHISTORY= Data Set

Variable	Description
Prin1–Prin$j$	Principal component scores
R_var1–R_$varp$	Residuals
NOBS	Number of observations used in the analysis
SPE	Squared prediction error (SPE)
TSQUARE	T^2 statistic computed from principal component scores

The names of the score variables are formed by concatenating the value given by the PREFIX= option (or the default Prin, if PREFIX= is not specified) and the numbers 1, 2, . . . , j, where j is the number of principal components in the model.

The names of the residual variables are formed by concatenating the value given by the RPREFIX= option (or the default R_, if RPREFIX= is not specified) and the names of the process variables used in the analysis. Residual variables are created only when the number of principal components in the model is less than the number of process measurement variables in the input data set.

OUTTABLE= Data Set

You can save control chart statistics, control limits, and related information by specifying the OUTTABLE= option in a SCORECHART, SPECHART, or TSQUARECHART statement. Each chart statement produces OUTTABLE= data sets containing different variables. Table 13.11, Table 13.12, and Table 13.13 list the variables that are contained in the three types of OUTTABLE= data set.

Table 13.11 Score Chart OUTTABLE= Data Set Variables

Variable	Description
COMP	Principal component number
EXLIM	Flag to indicate control limit was exceeded
LCL	Lower control limit
MEAN	Center line
SCORE	Principal component score
series	SERIES variable, if specified
SIGMAS	Multiple of score standard deviation used to compute control limits
time	TIME variable, if specified
UCL	Upper control limit

Table 13.12 SPE Chart OUTTABLE= Data Set Variables

Variable	Description
ALPHA	Probability (α) of exceeding control limits
EXLIM	Flag to indicate control limit was exceeded
LCL	Lower control limit
MEDIAN	Center line
NCOMP	Number of principal components in the model
residuals	Residual variables
series	SERIES variable, if specified
SPE	Squared prediction error (SPE) statistic
time	TIME variable, if specified
UCL	Upper control limit

Table 13.13 T^2 Chart OUTTABLE= Data Set Variables

Variable	Description
ALPHA	Probability (α) of exceeding control limits
EXLIM	Flag to indicate control limit was exceeded
LCL	Lower control limit
MEDIAN	Center line
NCOMP	Number of principal components in the model
residuals	Residual variables
series	SERIES variable, if specified
time	TIME variable, if specified
TSQUARE	T^2 statistic
UCL	Upper control limit

ODS Graphics

Before you create ODS Graphics output, ODS Graphics must be enabled (for example, by using the ODS GRAPHICS ON statement). For more information about enabling and disabling ODS Graphics, see the section "Enabling and Disabling ODS Graphics" (Chapter 21, *SAS/STAT User's Guide*).

The MVPMONITOR procedure assigns a name to each graph that it creates using ODS Graphics. You can use these names to refer to the graphs when you use ODS. The graph names are listed in Table 13.14.

Table 13.14 ODS Graphics Produced by PROC MVPMONITOR

ODS Graph Name	Plot Description	Statement or Option
ContributionPlot	Contribution plot	CONTRIBUTIONS option
ScoreChart	control chart of principal component scores	SCORECHART statement
SPEChart	Squared prediction error chart	SPECHART statement
TSquareChart	T^2 chart	TSQUARECHART statement

Examples: MVPMONITOR Procedure

The following examples use an airline flight delay data set similar to the one described in the section "Getting Started: MVPMONITOR Procedure" on page 952. The following statements create a data set named flightDelays, which contains average flight delays for each region of the continental United States:

```
data flightDelays;
   label flightDate='Date';
   format flightDate MMDDYY8.;
   input flightDate :MMDDYY8. region $ AA CO DL F9 FL NW UA US WN;
   datalines;
02/01/07 MW 14.9  7.1  7.9  8.5 14.8  4.5  5.1 13.4  5.1
02/01/07 NE 15.7  7.1  8.6  6.3 14.6  6.2  7.0 11.0  6.4
02/01/07 NW 17.8  2.6  6.1 28.8 11.6  6.1 11.6 27.3  3.7
02/01/07 SC 19.9  8.3 13.9  4.9 25.8 15.3  9.0 15.1 12.8
02/01/07 SE 16.1  1.9  8.7  8.7 15.1 18.3  4.0 10.4  6.5
02/01/07 SW 19.3  7.8  4.8 11.5 34.7  7.4  7.3 12.0  5.6
02/02/07 MW 14.3  9.6 14.1  6.2 12.8  6.0  3.9 15.3 11.4

   ... more lines ...

02/16/07 SE 29.4 13.5 16.8 19.7 11.4 10.4 27.7 34.2 20.8
02/16/07 SW 28.1 18.4 17.1 25.5  8.0 15.2 30.4 26.6 20.8
;
```

Example 13.1: Combining Data from Peer Processes

In some situations you might want to build a common principal component model by combining data from multiple peer processes that have similar patterns of stable variation. This enables you to borrow strength from the data. A common set of control limits is then computed for each peer process.

This example uses observations from all regions in the continental United States at each time value to construct a common principal component model. It then applies the model to flight data for one region.

The following statements create a principal component model that contains three principal components from the flightDelays data set and apply the model to data for the Midwest region:

```
proc mvpmodel data=flightDelays ncomp=3 noprint
            out=mvpair outloadings=mvpairloadings;
   var AA CO DL F9 FL NW UA US WN;
run;
proc mvpmonitor history=mvpair loadings=mvpairloadings;
   time flightDate;
   series region;
   spechart / seriesvalue='MW';
   tsquarechart / seriesvalue='MW';
run;
```

The flightDelays data set contains observations from all continental United States regions, with multiple observations (one for each region) at each time point as defined by the flightDate variable. The OUTLOADINGS=

data set that is produced by PROC MVPMODEL contains the model information. The SERIES statement specifies region as the variable that identifies sequences of related observations. The SERIESVALUE= option selects the Midwest region statistics to be plotted.

The resulting SPE chart is shown in Output 13.1.1.

Output 13.1.1 Multivariate Control Chart for SPE Statistics
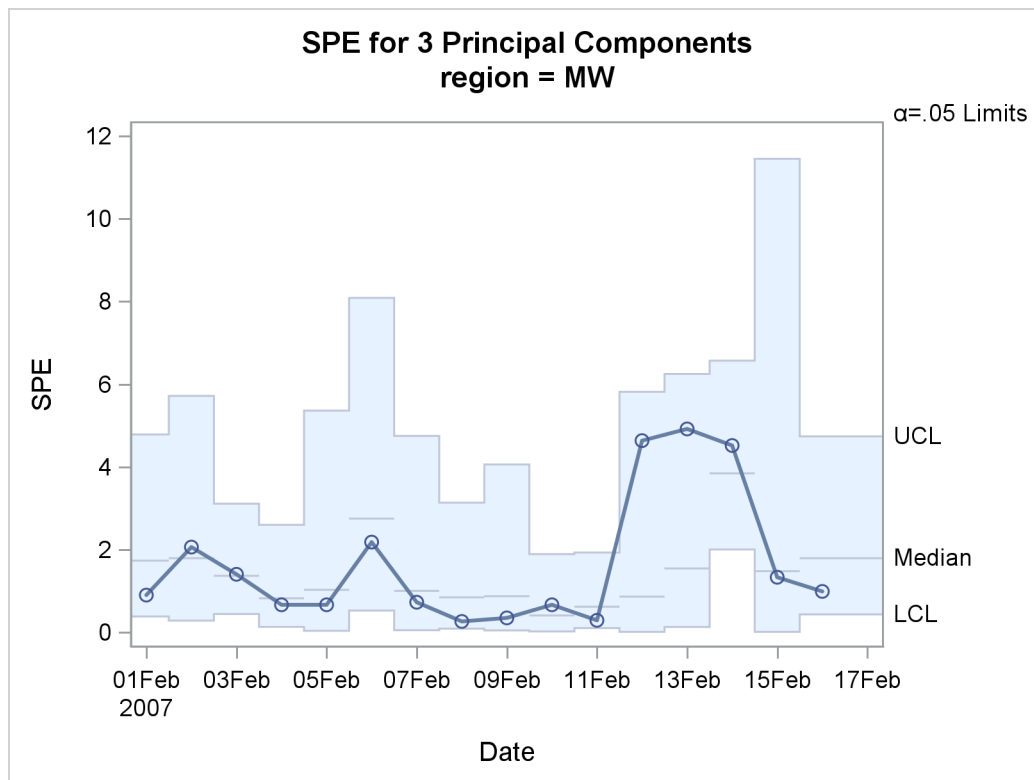
The control limits for the SPE chart are computed differently from a case with a single observation per time value, such as the chart shown in Figure 13.3. The control limits are based on different reference distributions for the SPE statistics in addition to different approximations to the reference distribution. See the section "Computing SPE Control Limits" on page 967 for more information.

The T^2 chart is shown in Output 13.1.2.

Output 13.1.2 Multivariate Control Chart for T^2 Statistics
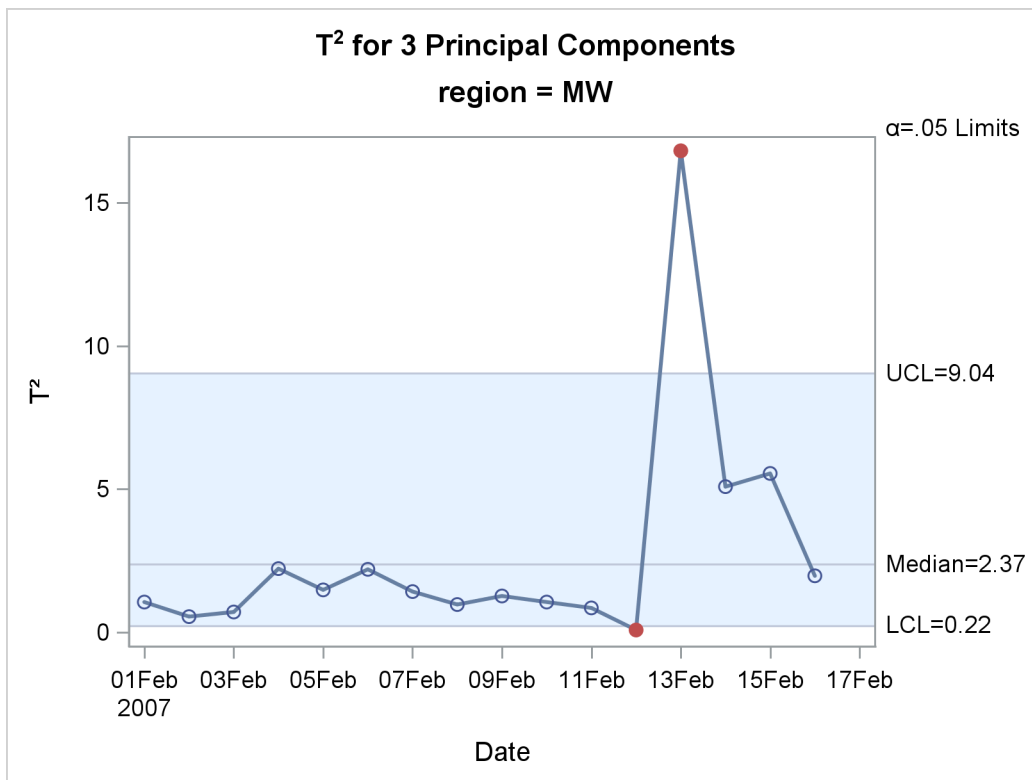
Compare the T^2 chart in Output 13.1.2 to the one in Figure 13.1. Both charts display T^2 statistics for the same flight delays from the Midwest region, but the charts are different because in this example the principal component model was constructed with data from all regions of the continental United States.

You can produce control charts for all the peer processes (regions in this example) by omitting the SERIES-VALUE= option. The following statements illustrate this approach:

```
proc mvpmonitor history=mvpair;
   time flightDate;
   series region;
   spechart;
   tsquarechart / overlay;
run;
```

By default, a separate control chart is created for each distinct value of the SERIES variable. The separate SPE charts for each region are not shown. The OVERLAY option in the TSQUARECHART statement specifies that the sequences for each region be plotted on a single T^2 chart, which is shown in Output 13.1.3.

Output 13.1.3 Overlaid T^2 Charts by Region
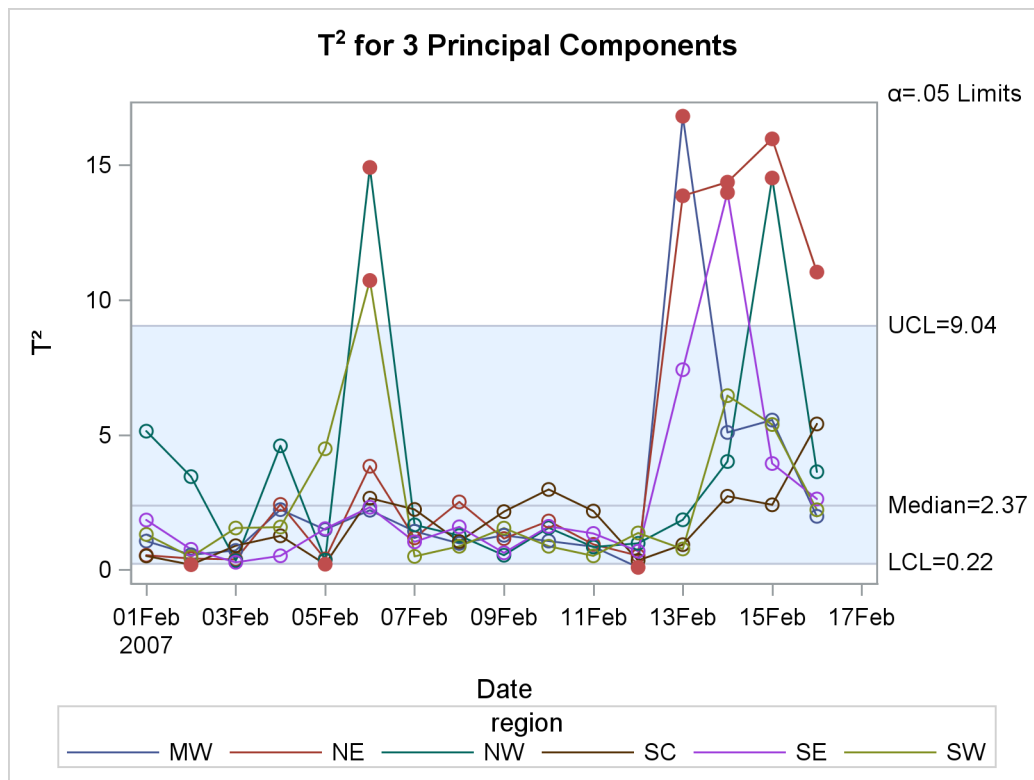
You can produce univariate control charts of standardized principal component scores by using the SCORECHART statement. The following statements produce control charts for the three principal components in the model and for each region:

```
proc mvpmonitor history=mvpair loadings=mvpairloadings;
   time flightDate;
   series region;
   scorechart / comp=all overlay=comp;
run;
```

The COMP=ALL option requests score charts for all the principal components in the model. The OVER-LAY=COMP option overlays the scores for each component in a single control chart. A separate chart is produced for each region. Output 13.1.4 and Output 13.1.5 show the score charts for the Midwest and Northeast regions, respectively.

Output 13.1.4 Score Charts for the Midwest Region
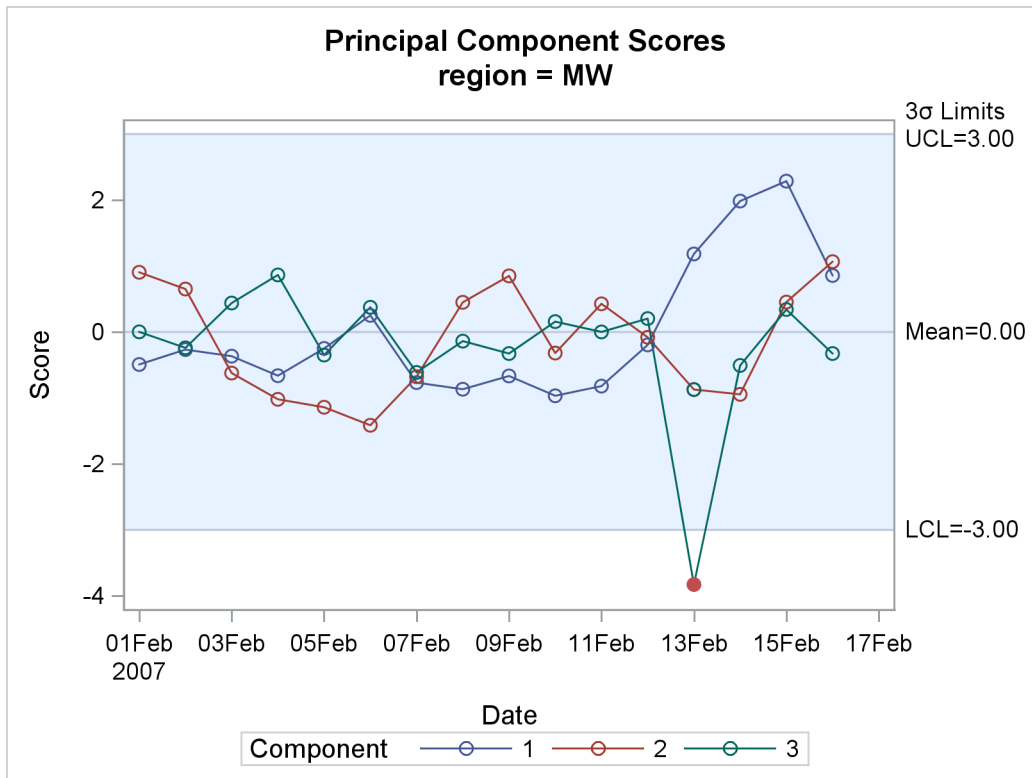
Output 13.1.5 Score Charts for the Northeast Region
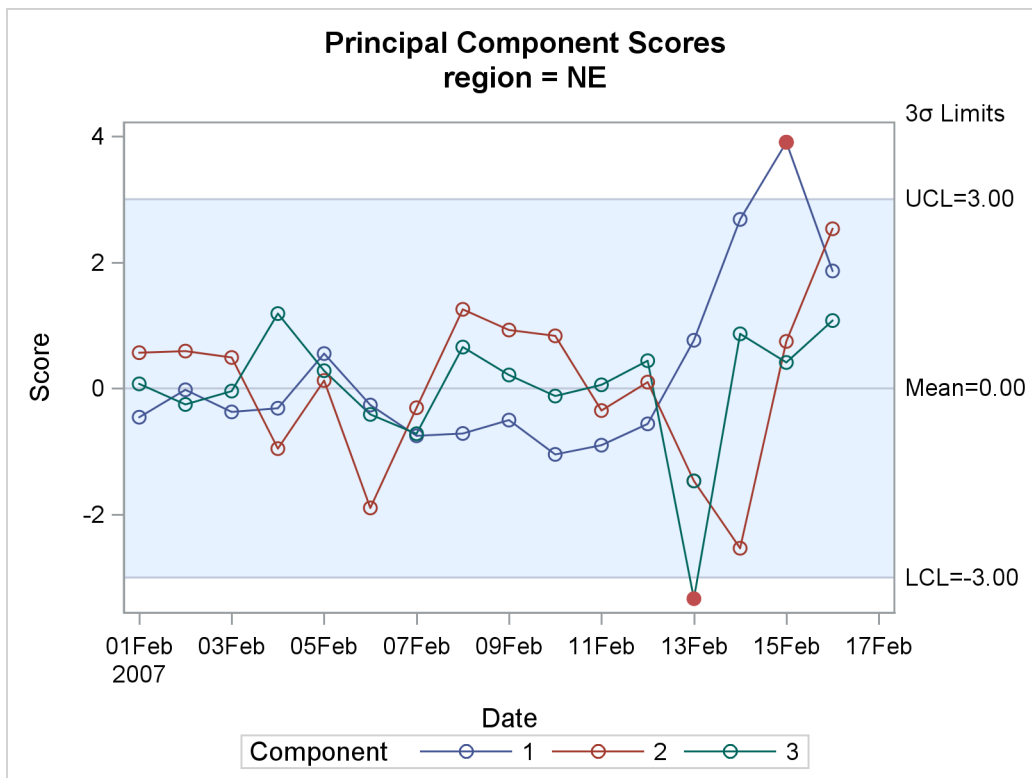
You can also overlay scores for different SERIES values in a single chart. The following statements produce a score chart for principal component 1 for each region:

```
proc mvpmonitor history=mvpair loadings=mvpairloadings;
   time flightDate;
   series region;
   scorechart / overlay=series;
run;
```

The resulting chart is shown in Output 13.1.6.

Output 13.1.6 Score Charts for Principal Component 1 and All Regions
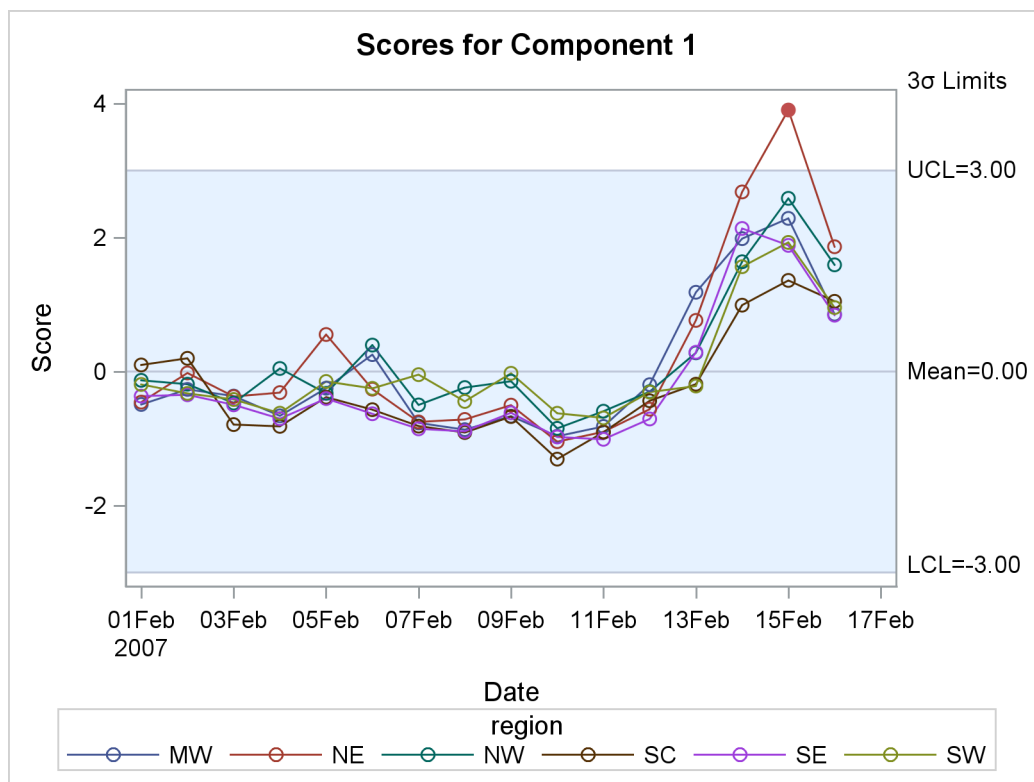
Example 13.2: Creating Multivariate Control Charts for Phase II

This example is a continuation of the example in "Getting Started: MVPMONITOR Procedure" on page 952. The following statements create a data set named flightDelays2, which provides flight delays for the date range February 17–28, 2007, for the northeastern United States:

```
data flightDelays2;
   label flightDate='Date';
   format flightDate MMDDYY8.;
   input flightDate :MMDDYY8. AA CO DL F9 FL NW UA US WN;
   datalines;
02/17/07 25.6  7.8 15.5 13.4 16.1 16.2 23.0 24.2  8.2
02/18/07  5.4 16.0  9.9  1.1 11.5 17.0 15.6 15.5  5.1
02/19/07 13.2 16.3 10.0 10.6  5.4 10.3  9.5 16.8  9.3
```

```
02/20/07  4.2   6.9   1.4   0.1   7.2   6.6   7.4 10.4   2.9
02/21/07  5.4  -0.1   7.4   8.7 16.3 24.3   9.4   6.0 10.2
02/22/07 19.6 30.2   6.8   2.7   8.9 16.4 14.3 12.6   8.2
02/23/07 14.9 18.9   9.9   9.1 12.0 16.5 17.4 12.8   6.0
02/24/07 21.4   5.5 11.1 46.1 10.6 55.3 22.9   8.8   3.4
02/25/07 42.6   7.7 14.6 14.4 32.0 50.7 46.1 49.4 39.1
02/26/07 43.2 25.1 18.1 18.2 28.8 31.1 38.6 29.6 18.6
02/27/07 11.3 17.1   5.3   4.1   4.8 13.9   9.8   9.7   7.1
02/28/07  8.1   3.7   2.7 17.1  -0.8   5.5 11.0 14.3   3.1
;
```

To use PROC MVPMONITOR in a Phase II analysis, you need a principal component model based on a process data from a stable process. The model that is produced by the MVPMODEL procedure in "Example 13.1: Combining Data from Peer Processes" on page 975 is used here. The model was generated from data for the continental United States during February 1–16, 2007. The model information is contained in the principal component loadings, which come from the mvpairloadings data set. The following statements apply the model to the new data in flightDelays2:

```
proc mvpmonitor data=flightDelays2 loadings=mvpairloadings;
    time flightDate;
    id flightDate;
    tsquarechart;
    spechart;
run;
```

The T^2 chart is shown in Output 13.2.1.

Output 13.2.1 Multivariate Control Chart for T^2 Statistics
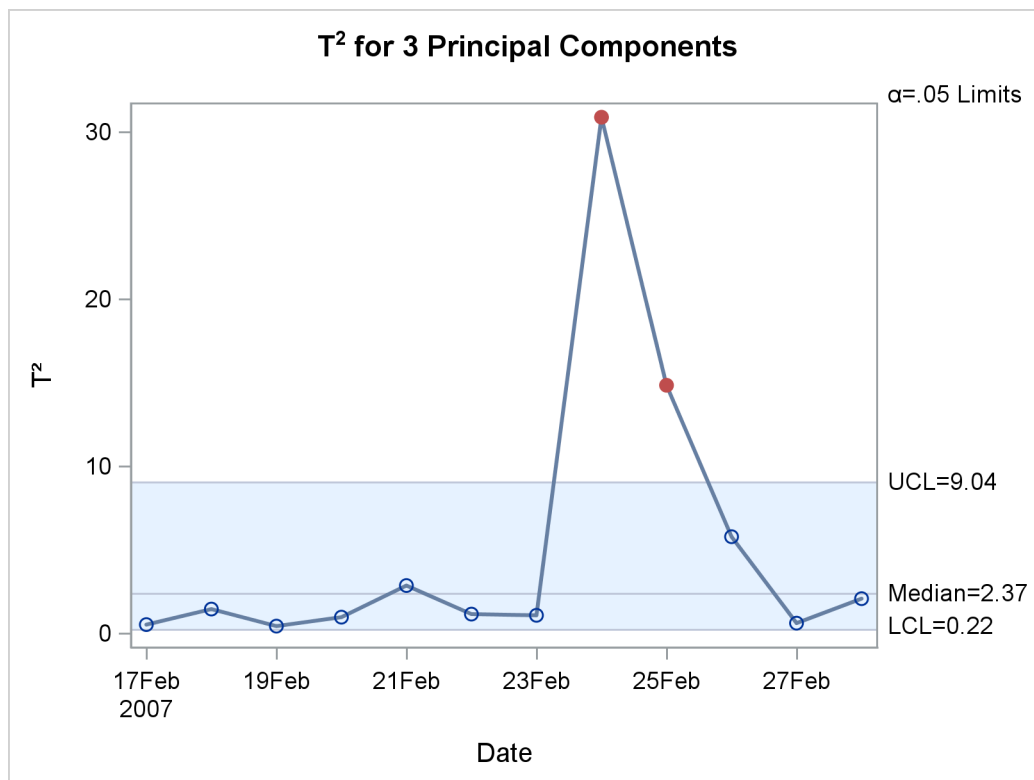
The SPE chart is shown in Output 13.2.2.

Output 13.2.2 Multivariate Control Chart for SPE Statistics
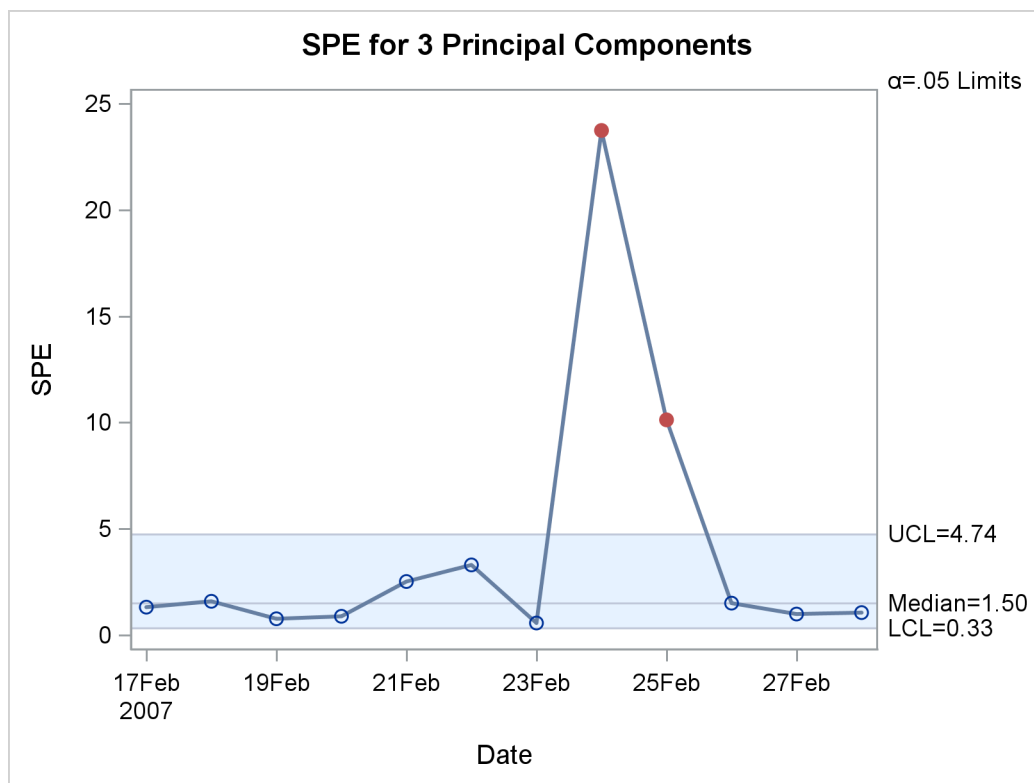
The SPE chart has out-of-control points on February 22, 24, and 25. This indicates that the variation is not along the model hyperplane, which suggests that the model is not appropriate for these new data.

Both the SPE chart and the T^2 chart have an out-of-control point on February 25. This point has very unusual variation. On that date, a major winter storm with high winds and blizzard conditions battered the Midwest while snow, sleet, and freezing rain hit the Northeast. These conditions contributed to delays that are not seen in the data set that the model was built with.

Example 13.3: Comparison of Univariate and Multivariate Control Charts

This example shows the effect of a change in correlation of the process variables on the SPE chart. The following statements create a data set called mvpStable, which consists of 30 samples from a trivariate normal distribution with strong positive correlation between all three variables:

```
proc iml;
   Mean = {0,0,0};
   Cor = {1.0 0.8 0.8,
          0.8 1.0 0.8,
          0.8 0.8 1.0};
   StdDevs = {2 2 2};
   D = diag(StdDevs);
   Cov = D*Cor*D; /* covariance matrix */
```

```
   NumSamples = 30;
   call randseed(123321);   /* set seed for the RandNormal module */
   X = RandNormal(NumSamples, Mean, Cov);
   varnames = { x1 x2 x3};
   create mvpStable from X [colname = varnames];
   append from X;
   quit;
run;
data mvpStable;
   set mvpStable;
   hour=_n_;
run;
```

The next statements create a data set called mvpOOC, which has five observations in which the correlations are negative:

```
proc iml;
   Mean = {0,0,0};
   Cor = { 1.0 -0.8  0.8,
          -0.8  1.0 -0.8,
           0.8 -0.8  1.0};
   StdDevs = {2 2 2};
   D = diag(StdDevs);
   Cov = D*Cor*D; /* covariance matrix */
   NumSamples = 5;
   call randseed(123321);   /* set seed for the RandNormal module */
   X = RandNormal(NumSamples, Mean, Cov);
   varnames = { x1 x2 x3};
   create mvpOOC from X [colname = varnames];
   append from X;
   quit;
run;
data mvpOOC;
   set mvpStable mvpOOC;
   hour=_n_;
run;
```

The following statements produce a principal component model for the data in mvpStable:

```
proc mvpmodel data=mvpStable ncomp=1 plots=none out=scores
              outloadings=loadings;
   var x1 x2 x3;
run;
```

The model hyperplane that is defined by specifying NCOMP= 1 is a line. The loadings in the principal component model, which are used to project the data to the model hyperplane, are defined by the correlation structure present in the DATA= data set.

The model explains about 90% of the variance in the data, as shown in Output 13.3.1.

Output 13.3.1 Eigenvalue and Variance Information

The MVPMODEL Procedure

| | Eigenvalues of the Correlation Matrix | | | |
	Eigenvalue	Difference	Proportion	Cumulative
1	2.67725690		0.8924	0.8924

The loadings from the model are then applied to the data in mvpOOC, which includes observations that have a different correlation structure, which vary in direction orthogonal to the model line. The following statements apply the loadings to these new data to produce T^2 and SPE charts:

```
proc mvpmonitor data=mvpOOC loadings=loadings;
    time hour;
    tsquarechart;
    spechart;
run;
```

The MVPMONITOR procedure generates a T^2 chart, shown in Output 13.3.2.

Output 13.3.2 T^2 Chart
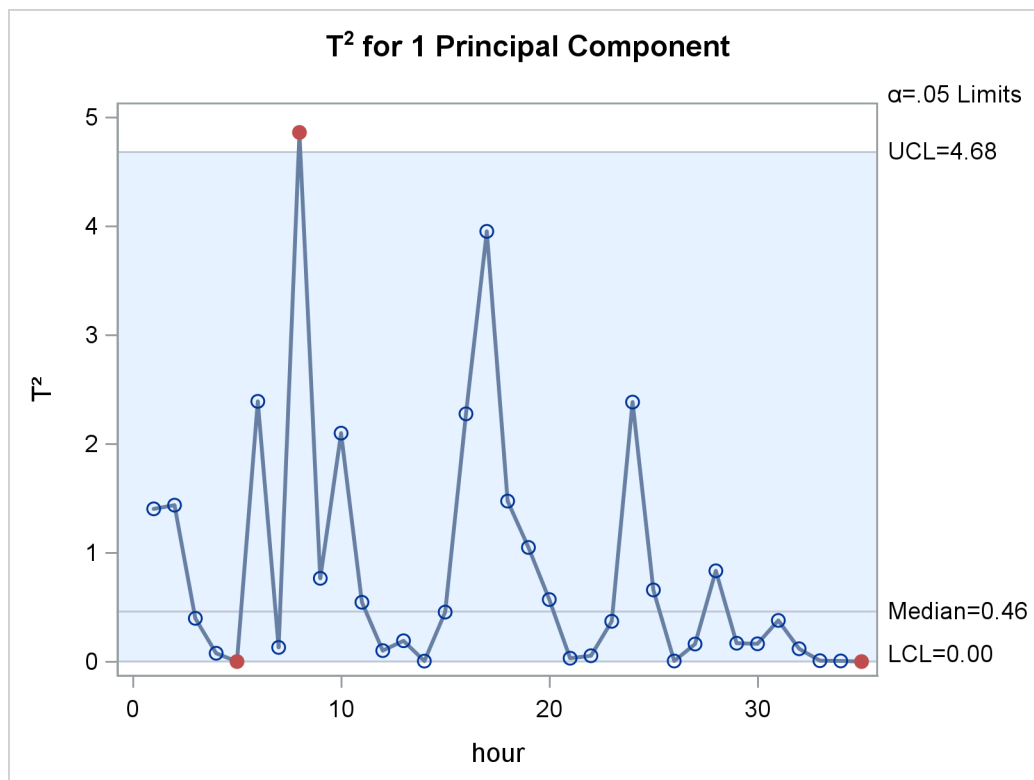
The projection of the last five points to the model line results in small amounts of variation, and thus small T^2 statistics, for two reasons: the last five points are orthogonal to the model line, and they share the same mean. However, the orthogonality means that they are out-of-control points in the SPE chart.

The MVPMONITOR procedure also produces an SPE chart, shown in Output 13.3.3.

Output 13.3.3 SPE Chart
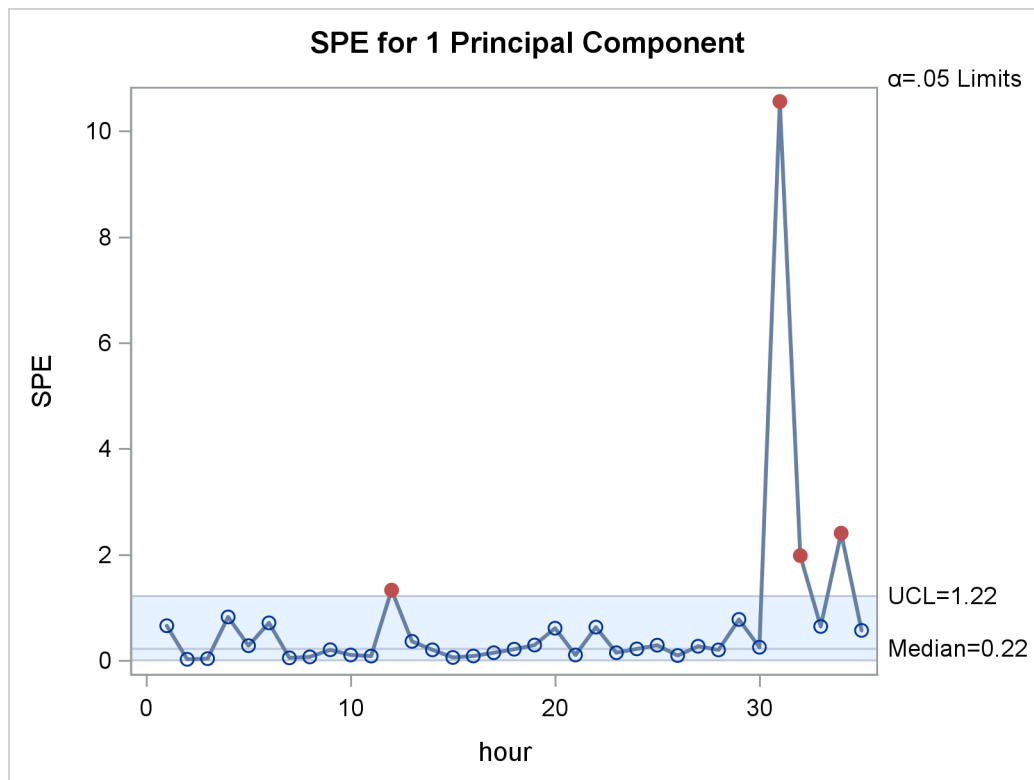
Because the last five points come from a correlation structure that is not seen in the data from which the model was built, these points can lie far from the model line, resulting in large values in the SPE statistics.

Because the marginal distributions are the same in both the original 30 points and the additional five points, the univariate control charts in Output 13.3.4, Output 13.3.5, and Output 13.3.6 fail to signal the multivariate change at hour 31. The following statements produce univariate control charts for each of the variables by using the SHEWHART procedure:

```
proc shewhart data=mvpOOC;
   irchart (x1 x2 x3) * hour / markers nochart2;
run;
```

Output 13.3.4 Univariate Chart for x_1
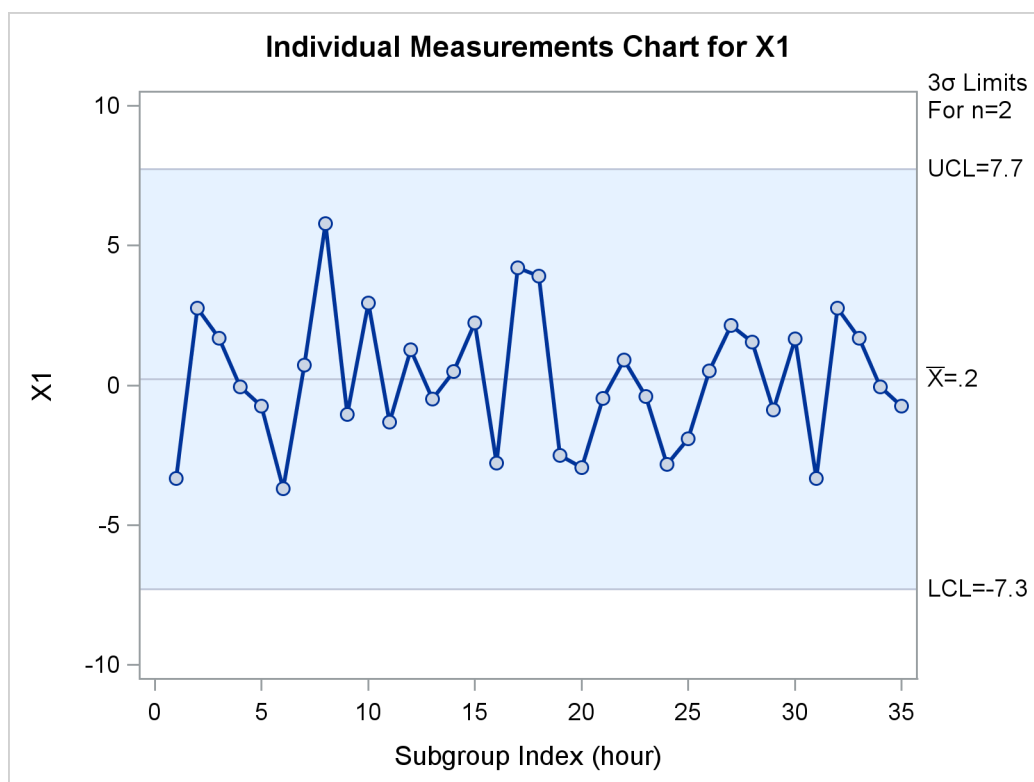
Output 13.3.5 Univariate Chart for x_2
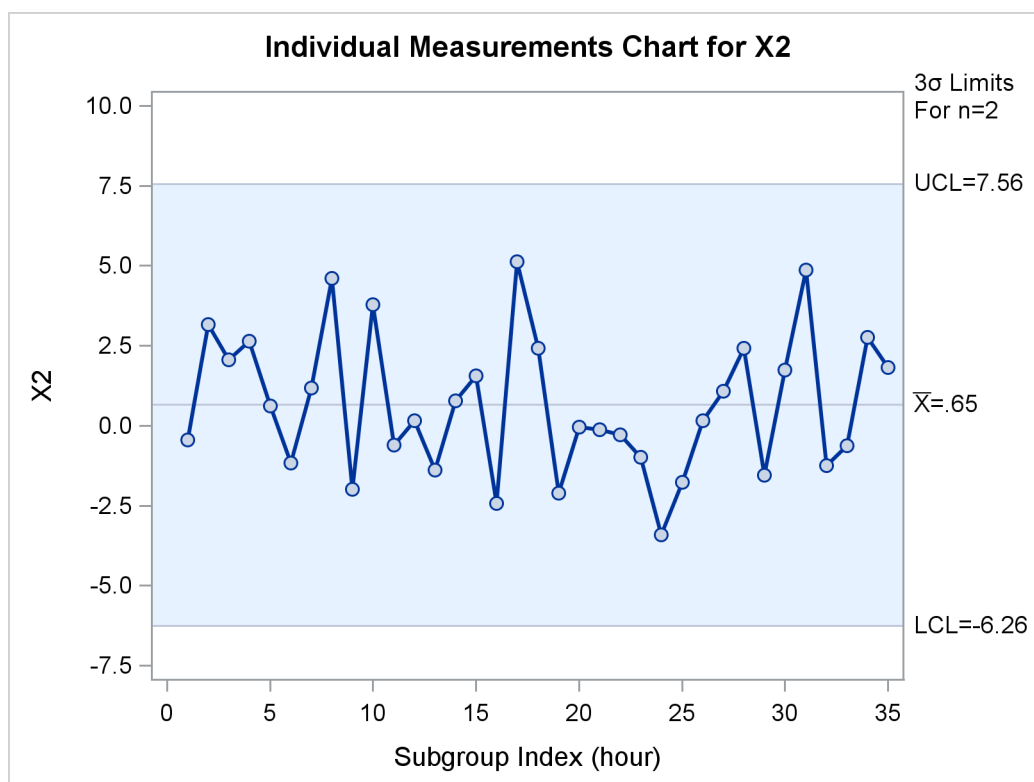
Output 13.3.6 Univariate Chart for x_3
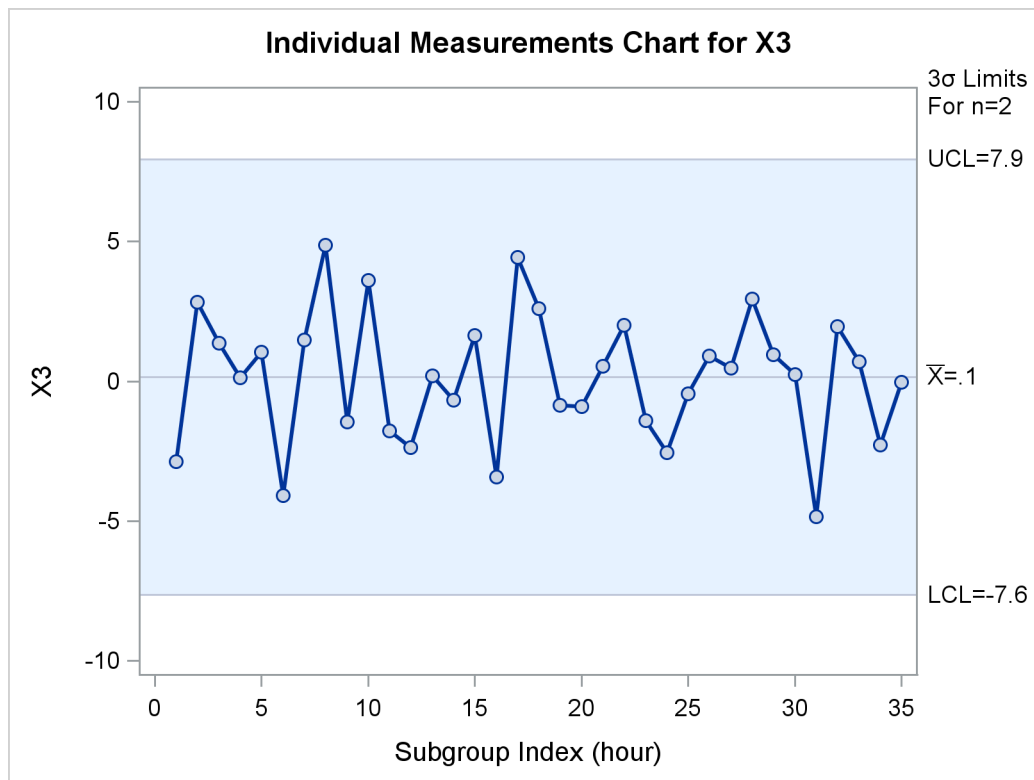
Example 13.4: Creating a Classical T^2 Chart

The following statements use PROC MVPMODEL to create a model from which classical T^2 charts can be produced:

```
proc mvpmodel data=flightDelays ncomp=all noprint out=mvpout;
   var AA CO DL F9 FL NW UA US WN;
run;
```

The NCOMP=ALL option specifies that the number of principal components equal the number of process variables, so the mvpout data set contains the classical T^2 statistic for each observation. The mvpout data set contains six observations per time point—one for each region.

The following statements create the classical T^2 chart:

```
proc mvpmonitor history=mvpout;
   time flightDate;
   series region;
   tsquarechart / seriesvalue='NE';
run;
```

The SERIES statement specifies region as the variable that identifies sequences of T^2 statistics, and the SERIESVALUE= option selects the region to be plotted, the Northeast. The classical T^2 chart is shown in Output 13.4.1.

Output 13.4.1 Classical T^2 Chart
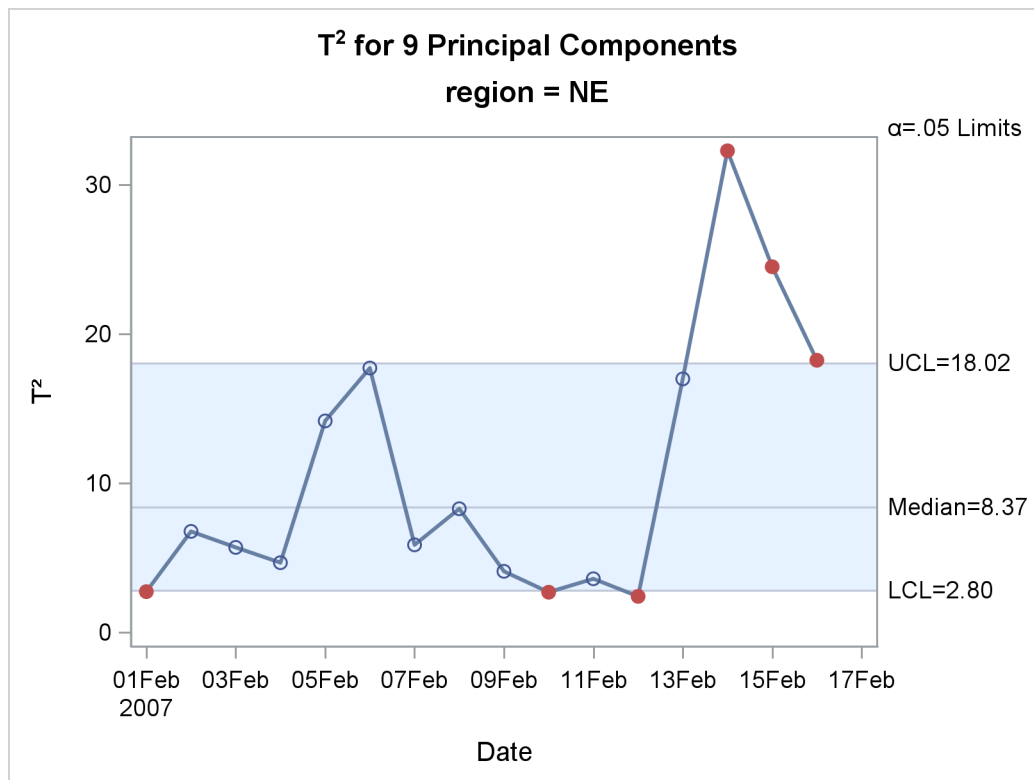
In this case, the classical T^2 chart finds out-of-control observations above the upper control limit during February 14–16 and below the lower control limit on February 1, 10, and 12.

Output 13.4.2 shows a partial listing of the mvpout data set. It contains T^2 statistics based on the model that has nine principal components, in addition to the original variables and other observationwise statistics.

Output 13.4.2 Partial Listing of Output Data Set mvpout

flightDate	region	AA	CO	DL	F9	FL	NW	UA	US	WN	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6
02/01/07	MW	14.9	7.1	7.9	8.5	14.8	4.5	5.1	13.4	5.1	-1.16440	0.89425	0.00118	-0.84496	0.07443	-0.05231
02/01/07	NE	15.7	7.1	8.6	6.3	14.6	6.2	7.0	11.0	6.4	-1.08197	0.55936	0.05859	-0.84487	-0.02092	-0.16851
02/01/07	NW	17.8	2.6	6.1	28.8	11.6	6.1	11.6	27.3	3.7	-0.31223	2.23593	-0.03824	0.58893	0.21628	0.67149
02/01/07	SC	19.9	8.3	13.9	4.9	25.8	15.3	9.0	15.1	12.8	0.22974	-0.16976	0.56118	-1.30139	0.22990	-0.34406
02/01/07	SE	16.1	1.9	8.7	8.7	15.1	18.3	4.0	10.4	6.5	-0.87056	0.01044	1.06975	-0.50179	0.38041	-0.33789

Prin7	Prin8	Prin9	_NOBS_	_TSQUARE_	R_AA	R_CO	R_DL	R_F9	R_FL	R_NW	R_UA	R_US	R_WN	_SPE_
-0.27403	0.31880	-0.06726	96	3.2122	0	0	0	0	0	0	0	0	0	.
-0.30596	0.25090	0.12804	96	2.7349	0	0	0	0	0	0	0	0	0	.
0.85708	1.07348	-0.49530	96	18.5555	0	0	0	0	0	0	0	0	0	.
0.11471	0.30201	0.46481	96	6.3335	0	0	0	0	0	0	0	0	0	.
0.63114	0.45028	0.21437	96	6.1537	0	0	0	0	0	0	0	0	0	.

Notice that no SPE statistics are produced when the number of principal components equals the number of process variables.

References

Alt, F. (1985). "Multivariate Quality Control." In *Encyclopedia of Statistical Sciences,* vol. 6, edited by S. Kotz, N. L. Johnson, and C. B. Read. New York: John Wiley & Sons.

Cooley, W. W., and Lohnes, P. R. (1971). *Multivariate Data Analysis*. New York: John Wiley & Sons.

Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations*. New York: John Wiley & Sons.

Hotelling, H. (1933). "Analysis of a Complex of Statistical Variables into Principal Components." *Journal of Educational Psychology* 24:417–441, 498–520.

Jackson, J. E., and Mudholkar, G. S. (1979). "Control Procedures for Residuals Associated with Principal Component Analysis." *Technometrics* 21:341–349.

Jensen, D. R., and Solomon, H. (1972). "A Gaussian Approximation to the Distribution of a Definite Quadratic Form." *Journal of the American Statistical Association* 67:898–902.

Kourti, T., and MacGregor, J. F. (1995). "Process Analysis, Monitoring and Diagnosis, Using Multivariate Projection Methods." *Chemometrics and Intelligent Laboratory Systems* 28:3–21.

Kourti, T., and MacGregor, J. F. (1996). "Multivariate SPC Methods for Process and Product Monitoring." *Journal of Quality Technology* 28:409–428.

Kshirsagar, A. M. (1972). *Multivariate Analysis*. New York: Marcel Dekker.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*. London: Academic Press.

Miller, P., Swanson, R. E., and Heckler, C. H. E. (1998). "Contribution Plots: A Missing Link in Multivariate Quality Control." *Applied Mathematics and Computer Science* 8:775–792.

Morrison, D. F. (1976). *Multivariate Statistical Methods*. 2nd ed. New York: McGraw-Hill.

Nomikos, P., and MacGregor, J. F. (1995). "Multivariate SPC Charts for Monitoring Batch Processes." *Technometrics* 37:41–59.

Pearson, K. (1901). "On Lines and Planes of Closest Fit to Systems of Points in Space." *Philosophical Magazine* 6:559–572.

Rao, C. R. (1964). "The Use and Interpretation of Principal Component Analysis in Applied Research." *Sankhyā, Series A* 26:329–358.

Tracy, N. D., Young, J. C., and Mason, R. L. (1992). "Multivariate Control Charts for Individual Observations." *Journal of Quality Technology* 24:88–95.

Wilks, S. S. (1962). *Mathematical Statistics*. New York: John Wiley & Sons.

Subject Index

Syntax Index